Data-X Fall 2018: Homework 8

Webscraping

Authors: Alexander Fred-Ojala

In this homework, you will do some exercises with web-scraping.

STUDENT NAME : JIEUN HWANG

SID :3033297165 ¶

Fun with Webscraping & Text manipulation

1. Statistics in Presidential Debates

Your first task is to scrape Presidential Debates from the Commission of Presidential Debates website: http://www.debates.org/index.php?page=debate-transcripts (http://www.debates.org/index.php?page=debate-transcripts).

To do this, you are not allowed to manually look up the URLs that you need, instead you have to scrape them. The root url to be scraped is the one listed above, namely: http://www.debates.org/index.php?page=debate-transcripts (http://www.debates.org/index.php?page=debate-transcripts)

1. By using `requests` and `BeautifulSoup` find all the links / URLs on the website that links to transcriptions of **First Presidential Debates** from the years [2012, 2008, 2004, 2000, 1996, 1988, 1984, 1976, 1960]. In total you should find 9 links / URLs tat fulfill this criteria. Print the urls.
2. When you have a list of the URLs your task is to create a Data Frame with some statistics (see example of output below):
   A. Scrape the title of each link and use that as the column name in your Data Frame.
   B. Count how long the transcript of the debate is (as in the number of characters in transcription string). Feel free to include `\` characters in your count, but remove any breakline characters, i.e. `\n`. You will get credit if your count is +/- 10% from our result.
   C. Count how many times the word **war** was used in the different debates. Note that you have to convert the text in a smart way (to not count the word **warranty** for example, but counting **war.**, **war!**, **war,** or **War** etc.
   D. Also scrape the most common used word in the debate, and write how many times it was used. Note that you have to use the same strategy as in 3 in order to do this.

   Print your final output result.

Tips:

In order to solve the questions above, it can be useful to work with Regular Expressions and explore methods on strings like `.strip(), .replace(), .find(), .count(), .lower()` etc. Both are very powerful tools to do string processing in Python. To count common words for example I used a `Counter` object and a Regular expression pattern for only words, see example:

```
from collections import Counter
    import re

    counts = Counter(re.findall(r"[\w']+", text.lower()))
```

Read more about Regular Expressions here: https://docs.python.org/3/howto/regex.html (https://docs.python.org/3/howto/regex.html)

Example output of all of the answers to Question 1.2:

	October 3, 2012: The First Obama-Romney Presidential Debate	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Debate char length	94627	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
war_count	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
most_common_w	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
most_common_w_count	757	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

.

In [59]:

```
import requests # The requests library is an
import bs4 as bs # BeautifulSoup4 is a Python library
import pandas as pd
from collections import Counter
import re
```

In [60]:

```
source = requests.get("http://www.debates.org/index.php?page=debate-transcripts")
soup = bs.BeautifulSoup(source.content, features='html.parser')
print(soup.body)
```

```
<body>
<div id="wrapper">
<div id="header">
</div>
<div id="menubar">
<a href="http://www.debates.org/" id="tm-1">Home</a>
<a href="http://www.debates.org/index.php?page=about-cpd" id="tm-2">About CPD</a>
<a href="http://www.debates.org/index.php?page=debate-history" id="tm-3">Debate History</a>
<a href="http://www.debates.org/index.php?page=news" id="tm-4">News</a>
<a class="current" href="http://www.debates.org/index.php?page=voter-education" id="tm-5">Voter Educati
```

```
on</a>
<a href="http://www.debates.org/index.php?page=international" id="tm-6">International</a>
<a href="http://www.debates.org/index.php?page=media" id="tm-11">Media</a>
</div>
<div id="searchbar">
<div id="searchinput">
<form action="http://www.debates.org/index.php?page=search-results" class="cms_form" id="cntnt01modulef
orm_1" method="get">
<div class="hidden">
<input name="mact" type="hidden" value="Search,cntnt01,dosearch,0"/>
<input name="cntnt01returnid" type="hidden" value="36"/>
</div>
<input id="cntnt01searchinput" maxlength="50" name="cntnt01searchinput" onblur="if(this.value=='') this
.value=this.defaultValue;" onfocus="if(this.value==this.defaultValue) this.value='';" size="20" type="t
ext" value="Enter Search..."/><a href="#" id="imgSearch" onclick="document.forms[0].submit();"></a>
<input id="cntnt01origreturnid" name="cntnt01origreturnid" type="hidden" value="39"/>
</form>
</div>
<div style="clear: both"></div>
</div>
<div id="printf">
</div>
<div style="clear: both;"></div>
<div id="contentwrapper">
<div id="leftmenu">
<span class="lmtop">In This Section</span>
<hr/>
<ul class="arrow1">
<li><a href="http://www.debates.org/index.php?page=voter-education"> Voter Education Overview </a>
</li>
<li><a href="http://www.debates.org/index.php?page=guide-to-hosting-your-own-debate"> Guide to Hosting
Your Own Debate </a>
</li>
<li><a href="http://www.debates.org/index.php?page=debatewatch-overview"> Host a DebateWatch </a>
</li>
<li><a href="http://www.debates.org/index.php?page=voter-education-partners"> Voter Education Partners
</a>
<ul>
<li><a href="http://www.debates.org/index.php?page=debatewatch-2000-organizational-partners"> DebateWat
ch 2000 Organizational Partners </a>
</li>
<li><a href="http://www.debates.org/index.php?page=organizations-participating-in-debatewatch-2004"> Or
ganizations Participating in DebateWatch 2004 </a>
</li>
<li><a href="http://www.debates.org/index.php?page=debatewatch-2000-academic-partners"> DebateWatch 200
0 Academic Partners </a>
</li></ul>
</li>
<li><a href="http://www.debates.org/index.php?page=citizen-resources"> Citizen Resources </a>
</li>
<li><a href="http://www.debates.org/index.php?page=debate-videos"> Debate Videos </a>
</li>
<li>Debate Transcripts
<ul>
<li><a href="http://www.debates.org/index.php?page=october-19-2016-debate-transcript"> October 19, 2016
Debate Transcript </a>
</li>
<li><a href="http://www.debates.org/index.php?page=october-9-2016-debate-transcript"> October 9, 2016 D
ebate Transcript </a>
</li>
<li><a href="http://www.debates.org/index.php?page=october-4-2016-debate-transcript"> October 4, 2016 D
ebate Transcript </a>
</li>
<li><a href="http://www.debates.org/index.php?page=september-26-2016-debate-transcript"> September 26,
2016 Debate Transcript </a>
</li>
<li><a href="http://www.debates.org/index.php?page=october-22-2012-the-third-obama-romney-presidential-
debate"> October 22, 2012 Debate Transcript </a>
</li>
<li><a href="http://www.debates.org/index.php?page=october-16-2012-the-second-obama-romney-presidential
-debate"> October 16, 2012 Debate Transcript </a>
</li>
<li><a href="http://www.debates.org/index.php?page=october-11-2012-the-biden-romney-vice-presidential-d
ebate"> October 11, 2012 Debate Transcript </a>
</li>
<li><a href="http://www.debates.org/index.php?page=october-3-2012-debate-transcript"> October 3, 2012 D
ebate Transcript </a>
</li>
<li><a href="http://www.debates.org/index.php?page=2008-debate-transcript"> September 26, 2008 Debate T
```

```
ranscript </a>
</li>
<li><a href="http://www.debates.org/index.php?page=2008-debate-transcript-2"> October 2, 2008 Debate Tr
anscript </a>
</li>
<li><a href="http://www.debates.org/index.php?page=october-7-2008-debate-transcrip"> October 7, 2008 De
bate Transcrip </a>
</li>
<li><a href="http://www.debates.org/index.php?page=october-15-2008-debate-transcript"> October 15, 2008
Debate Transcript </a>
</li>
<li><a href="http://www.debates.org/index.php?page=october-13-2004-debate-transcript"> October 13, 2004
Debate Transcript </a>
</li>
<li><a href="http://www.debates.org/index.php?page=october-8-2004-debate-transcript"> October 8, 2004 D
ebate Transcript </a>
</li>
<li><a href="http://www.debates.org/index.php?page=october-5-2004-transcript"> October 5, 2004 Transcri
pt </a>
</li>
<li><a href="http://www.debates.org/index.php?page=september-30-2004-debate-transcript"> September 30.
2004 Debate Transcript </a>
</li>
<li><a href="http://www.debates.org/index.php?page=october-3-2000-transcript"> October 3, 2000 Transcri
pt </a>
</li>
<li><a href="http://www.debates.org/index.php?page=october-5-2000-debate-transcript"> October 5, 2000 D
ebate Transcript </a>
</li>
<li><a href="http://www.debates.org/index.php?page=october-11-2000-debate-transcript"> October 11, 2000
Debate Transcript </a>
</li>
<li><a href="http://www.debates.org/index.php?page=october-17-2000-debate-transcript"> October 17, 2000
Debate Transcript </a>
</li>
<li><a href="http://www.debates.org/index.php?page=october-6-1996-debate-transcript"> October 6, 1996 D
ebate Transcript </a>
</li>
<li><a href="http://www.debates.org/index.php?page=october-9-1996-debate-transcript"> October 9, 1996 D
ebate Transcript </a>
</li>
<li><a href="http://www.debates.org/index.php?page=october-16-1996-debate-transcript"> October 16, 1996
Debate Transcript </a>
</li>
<li><a href="http://www.debates.org/index.php?page=october-11-1992-first-half-debate-transcript"> Octob
er 11, 1992 First Half Debate Transcript </a>
</li>
<li><a href="http://www.debates.org/index.php?page=october-11-1992-second-half-debate-transcript"> Octo
ber 11, 1992 Second Half Debate Transcript </a>
</li>
<li><a href="http://www.debates.org/index.php?page=october-13-1992-debate-transcript"> October 13, 1992
Debate Transcript </a>
</li>
<li><a href="http://www.debates.org/index.php?page=october-15-1992-first-half-debate-transcript"> Octob
er 15, 1992 First Half Debate Transcript </a>
</li>
<li><a href="http://www.debates.org/index.php?page=october-15-1992-second-half-debate-transcript"> Octo
ber 15, 1992 Second Half Debate Transcript </a>
</li>
<li><a href="http://www.debates.org/index.php?page=october-19-1992-debate-transcript"> October 19, 1992
Debate Transcript </a>
</li>
<li><a href="http://www.debates.org/index.php?page=september-25-1988-debate-transcript"> September 25,
1988 Debate Transcript </a>
</li>
<li><a href="http://www.debates.org/index.php?page=october-5-1988-debate-transcripts"> October 5, 1988
Debate Transcripts </a>
</li>
<li><a href="http://www.debates.org/index.php?page=october-13-1988-debate-transcript"> October 13, 1988
Debate Transcript </a>
</li>
<li><a href="http://www.debates.org/index.php?page=october-7-1984-debate-transcript"> October 7, 1984 D
ebate Transcript </a>
</li>
<li><a href="http://www.debates.org/index.php?page=october-11-1984-debate-transcript"> October 11, 1984
Debate Transcript </a>
</li>
<li><a href="http://www.debates.org/index.php?page=october-21-1984-debate-transcript"> October 21, 1984
Debate Transcript </a>
```

```
</li>
<li><a href="http://www.debates.org/index.php?page=september-21-1980-debate-transcript"> September 21,
1980 Debate Transcript </a>
</li>
<li><a href="http://www.debates.org/index.php?page=october-28-1980-debate-transcript"> October 28, 1980
Debate Transcript </a>
</li>
<li><a href="http://www.debates.org/index.php?page=september-23-1976-debate-transcript"> September 23,
1976 Debate Transcript </a>
</li>
<li><a href="http://www.debates.org/index.php?page=october-6-1976-debate-transcript"> October 6, 1976 D
ebate Transcript </a>
</li>
<li><a href="http://www.debates.org/index.php?page=october-22-1976-debate-transcript"> October 22, 1976
Debate Transcript </a>
</li>
<li><a href="http://www.debates.org/index.php?page=september-26-1960-debate-transcript"> September 26,
1960 Debate Transcript </a>
</li>
<li><a href="http://www.debates.org/index.php?page=october-7-1960-debate-transcript"> October 7, 1960 D
ebate Transcript </a>
</li>
<li><a href="http://www.debates.org/index.php?page=october-13-1960-debate-transcript"> October 13, 1960
Debate Transcript </a>
</li>
<li><a href="http://www.debates.org/index.php?page=october-21-1960-debate-transcript"> October 21, 1960
Debate Transcript </a>
</li>
<li><a href="http://www.debates.org/index.php?page=2000-debate-transcripts-translations"> Debate Transc
ript Translations </a>
</li></ul>
</li>
<li><a href="http://www.debates.org/index.php?page=state-boards-of-election"> State Boards of Election
</a>
</li>
<li><a href="http://www.debates.org/index.php?page=tips-for-organizations-schools-or-students"> Tips fo
r Organizations, Schools, or Students </a>
</li>
</ul>
</div>
<div id="content-sm">
<h1>Debate Transcripts</h1>
<p>
<span class="graytext"><span class="grayheading">Unofficial transcripts of most presidential and vice p
residential debates are available on this site.</span><br/><br/></span>
<hr/>
<p class="graytext">2012 Transcipts</p>
<blockquote>
<p><a href="http://www.debates.org/index.php?page=october-3-2012-debate-transcript" title="October 3, 2
012 Debate Transcript">October 3, 2012: The First Obama-Romney Presidential Debate</a></p>
<p><a href="index.php?page=october-11-2012-the-biden-romney-vice-presidential-debate" target="_blank">O
ctober 11, 2012: The Biden-Ryan Vice Presidential Debate</a></p>
<p><a href="index.php?page=october-1-2012-the-second-obama-romney-presidential-debate" target="_blank">
October 16, 2012: The Second Obama-Romney Presidential Debate</a></p>
<p><a href="index.php?page=october-22-2012-the-third-obama-romney-presidential-debate" target="_blank">
October 22, 2012: The Third Obama-Romney Presidential Debate</a></p>
</blockquote>
<hr/>
<p class="graytext">2008 Transcripts</p>
<blockquote>
<p><a href="http://www.debates.org/index.php?page=2008-debate-transcript" title="September 26, 2008 Deb
ate Transcript">September 26, 2008: The First McCain-Obama Presidential Debate</a></p>
<p><a href="http://www.debates.org/index.php?page=2008-debate-transcript-2" title="October 2, 2008 Deba
te Transcript">October 2, 2008: The Biden-Palin Vice Presidential Debate</a></p>
<p><a href="http://www.debates.org/index.php?page=october-7-2008-debate-transcrip" title="October 7, 20
08 Debate Transcript">October 7, 2008: The Second McCain-Obama Presidential Debate</a></p>
<p><a href="http://www.debates.org/index.php?page=october-15-2008-debate-transcript" title="October 15,
2008 Debate Transcript">October 15, 2008: The Third McCain-Obama Presidential Debate</a></p>
</blockquote>
<hr/>
<p class="graytext">2004 Transcripts</p>
<blockquote>
<p><a href="http://www.debates.org/index.php?page=october-13-2004-debate-transcript" title="October 13,
2004 Debate Transcript">October 13, 2004: The Third Bush-Kerry Presidential Debate</a></p>
<p><a href="http://www.debates.org/index.php?page=october-8-2004-debate-transcript" title="October 8, 2
004 Debate Transcript">October 8, 2004: The Second Bush-Kerry Presidential Debate</a></p>
<p><a href="http://www.debates.org/index.php?page=october-5-2004-transcript" title="October 5, 2004 Tra
nscript">October 5, 2004: The Cheney-Edwards Vice Presidential Debate</a></p>
<p><a href="http://www.debates.org/index.php?page=september-30-2004-debate-transcript" title="September
```

```
30. 2004 Debate Transcript">September 30, 2004: The First Bush-Kerry Presidential Debate</a></p>
</blockquote>
<hr/>
<p class="graytext">2000 Transcripts</p>
<blockquote>
<p><a href="http://www.debates.org/index.php?page=october-3-2000-transcript" title="October 3, 2000 Tra
nscript">October 3, 2000: The First Gore-Bush Presidential Debate</a></p>
<p><a href="http://www.debates.org/index.php?page=october-5-2000-debate-transcript" title="October 5, 2
000 Debate Transcript">October 5, 2000: The Lieberman-Cheney Vice Presidential Debate</a></p>
<p><a href="http://www.debates.org/index.php?page=october-11-2000-debate-transcript" title="October 11,
2000 Debate Transcript">October 11, 2000: The Second Gore-Bush Presidential Debate</a></p>
<p><a href="http://www.debates.org/index.php?page=october-17-2000-debate-transcript" title="October 17,
2000 Debate Transcript">October 17, 2000: The Third Gore-Bush Presidential Debate</a></p>
<p><a href="http://www.debates.org/index.php?page=2000-debate-transcripts-translations" title="Debate T
ranscript Translations">The 2000 Debate Transcripts: Transcripts of the debates translated into six lan
guages</a></p>
</blockquote>
<hr/>
<p class="graytext">1996 Transcripts</p>
<blockquote>
<p><a href="http://www.debates.org/index.php?page=october-6-1996-debate-transcript" title="October 6, 1
996 Debate Transcript">October 6, 1996: The First Clinton-Dole Presidential Debate</a></p>
<p><a href="http://www.debates.org/index.php?page=october-9-1996-debate-transcript" title="October 9, 1
996 Debate Transcript">October 9, 1996: The Gore-Kemp Vice Presidential Debate</a></p>
<p><a href="http://www.debates.org/index.php?page=october-16-1996-debate-transcript" title="October 16,
1996 Debate Transcript">October 16, 1996: The Second Clinton-Dole Presidential Debate</a></p>
</blockquote>
<hr/>
<p class="graytext">1992 Transcripts</p>
<blockquote>
<p>October 11, 1992: The First Clinton-Bush-Perot Presidential Debate</p>
<p><a href="http://www.debates.org/index.php?page=october-11-1992-first-half-debate-transcript" title="
October 11, 1992 First Half Debate Transcript">First half of Debate</a><br/><a href="http://www.debates
.org/index.php?page=october-11-1992-second-half-debate-transcript" title="October 11, 1992 Second Half
Debate Transcript">Second half of Debate</a></p>
<p><a href="http://www.debates.org/index.php?page=october-13-1992-debate-transcript" title="October 13,
1992 Debate Transcript">October 13, 1992: The Gore-Quayle-Stockdale Vice Presidential Debate</a></p>
<p>October 15, 1992: The Second Clinton-Bush-Perot Presidential Debate</p>
<p><a href="http://www.debates.org/index.php?page=october-15-1992-first-half-debate-transcript" title="
October 15, 1992 First Half Debate Transcript">First half of Debate</a><br/><a href="http://www.debates
.org/index.php?page=october-15-1992-second-half-debate-transcript" title="October 15, 1992 Second Half
Debate Transcript">Second half of Debate</a></p>
<p><a href="http://www.debates.org/index.php?page=october-19-1992-debate-transcript" title="October 19,
1992 Debate Transcript">October 19, 1992: The Third Clinton-Bush-Perot Presidential Debate</a></p>
</blockquote>
<hr/>
<p class="graytext">1988 Transcripts</p>
<blockquote>
<p><a href="http://www.debates.org/index.php?page=september-25-1988-debate-transcript" title="September
25, 1988 Debate Transcript">September 25, 1988: The First Bush-Dukakis Presidential Debate</a></p>
<p><a href="http://www.debates.org/index.php?page=october-5-1988-debate-transcripts" title="October 5,
1988 Debate Transcripts">October 5, 1988: The Bentsen-Quayle Vice Presidential Debate</a></p>
<p><a href="http://www.debates.org/index.php?page=october-13-1988-debate-transcript" title="October 13,
1988 Debate Transcript">October 13, 1988: The Second Bush-Dukakis Presidential Debate</a></p>
</blockquote>
<hr/>
<p class="graytext">1984 Transcripts</p>
<blockquote>
<p><a href="http://www.debates.org/index.php?page=october-7-1984-debate-transcript" title="October 7, 1
984 Debate Transcript">October 7, 1984: The First Reagan-Mondale Presidential Debate</a></p>
<p><a href="http://www.debates.org/index.php?page=october-11-1984-debate-transcript" title="October 11,
1984 Debate Transcript">October 11, 1984: The Bush-Ferraro Vice Presidential Debate</a></p>
<p><a href="http://www.debates.org/index.php?page=october-21-1984-debate-transcript" title="October 21,
1984 Debate Transcript">October 21, 1984: The Second Reagan-Mondale Presidential Debate</a></p>
</blockquote>
<hr/>
<p class="graytext">1980 Transcripts</p>
<blockquote>
<p><a href="http://www.debates.org/index.php?page=september-21-1980-debate-transcript" title="September
21, 1980 Debate Transcript">September 21, 1980: The Anderson-Reagan Presidential Debate</a></p>
<p><a href="http://www.debates.org/index.php?page=october-28-1980-debate-transcript" title="October 28,
1980 Debate Transcript">October 28, 1980: The Carter-Reagan Presidential Debate</a></p>
</blockquote>
<hr/>
<p class="graytext">1976 Transcripts</p>
<blockquote><dl><dt><a href="http://www.debates.org/index.php?page=september-23-1976-debate-transcript"
title="September 23, 1976 Debate Transcript">September 23, 1976: The First Carter-Ford Presidential Deb
ate</a></dt><dt><br/></dt><dt><a href="http://www.debates.org/index.php?page=october-6-1976-debate-tran
script" title="October 6, 1976 Debate Transcript">October 6, 1976: The Second Carter-Ford Presidential
```

```
Debate</a></dt><dt><br/></dt><dt><a href="http://www.debates.org/index.php?page=october-22-1976-debate-
transcript" title="October 22, 1976 Debate Transcript">October 22, 1976: The Third Carter-Ford Presiden
tial Debate</a></dt></dl></blockquote>
<hr/>
<p class="graytext">1960 Transcripts</p>
<blockquote>
<p><a href="http://www.debates.org/index.php?page=september-26-1960-debate-transcript" title="September
26, 1960 Debate Transcript">September 26, 1960: The First Kennedy-Nixon Presidential Debate</a></p>
<p><a href="http://www.debates.org/index.php?page=october-7-1960-debate-transcript" title="October 7, 1
960 Debate Transcript">October 7, 1960: The Second Kennedy-Nixon Presidential Debate</a></p>
<p><a href="http://www.debates.org/index.php?page=october-13-1960-debate-transcript" title="October 13,
1960 Debate Transcript">October 13, 1960: The Third Kennedy-Nixon Presidential Debate</a></p>
<p><a href="http://www.debates.org/index.php?page=october-21-1960-debate-transcript" title="October 21,
1960 Debate Transcript">October 21, 1960: The Fourth Kennedy-Nixon Presidential Debate</a></p>
</blockquote>
<br/>
</p>
</div>
</div>
<div style="clear: both"></div>
<div id="footer">
<p>© COPYRIGHT 2015 THE COMMISSION ON PRESIDENTIAL DEBATES. ALL RIGHTS RESERVED.</p>
<script type="text/javascript">

  var _gaq = _gaq || [];
  _gaq.push(['_setAccount', 'UA-33437117-1']);
  _gaq.push(['_setDomainName', 'debates.org']);
  _gaq.push(['_trackPageview']);

  (function() {
    var ga = document.createElement('script'); ga.type = 'text/javascript'; ga.async = true;
    ga.src = ('https:' == document.location.protocol ? 'https://ssl' : 'http://www') + '.google-analyti
cs.com/ga.js';
    var s = document.getElementsByTagName('script')[0]; s.parentNode.insertBefore(ga, s);
  })();

</script>
</div>
<div id="bottommenu" style="display:none;">
<a href="http://www.debates.org/">Home</a>
  |
<a href="http://www.debates.org/index.php?page=about-cpd">About CPD</a>
  |
<a href="http://www.debates.org/index.php?page=debate-history">Debate History</a>
  |
<a href="http://www.debates.org/index.php?page=news">News</a>
  |
<a href="http://www.debates.org/index.php?page=voter-education">Voter Education</a>
  |
<a href="http://www.debates.org/index.php?page=international">International</a>
  |
<a href="http://www.debates.org/index.php?page=media">Media</a>
</div>
</div>
</body>
```

In [61]:
```
Debates = soup.find(id="content-sm")
print(Debates.prettify())
```

```
<div id="content-sm">
 <h1>
  Debate Transcripts
 </h1>
 <p>
  <span class="graytext">
   <span class="grayheading">
    Unofficial transcripts of most presidential and vice presidential debates are available on this sit
e.
   </span>
   <br/>
   <br/>
  </span>
  <hr/>
  <p class="graytext">
   2012 Transcipts
  </p>
  <blockquote>
   <p>
    <a href="http://www.debates.org/index.php?page=october-3-2012-debate-transcript" title="October 3,
2012 Debate Transcript">
     October 3, 2012: The First Obama-Romney Presidential Debate
    </a>
   </p>
   <p>
    <a href="index.php?page=october-11-2012-the-biden-romney-vice-presidential-debate" target="_blank">
     October 11, 2012: The Biden-Ryan Vice Presidential Debate
    </a>
   </p>
   <p>
    <a href="index.php?page=october-1-2012-the-second-obama-romney-presidential-debate" target="_blank"
>
     October 16, 2012: The Second Obama-Romney Presidential Debate
    </a>
   </p>
   <p>
    <a href="index.php?page=october-22-2012-the-third-obama-romney-presidential-debate" target="_blank"
>
     October 22, 2012: The Third Obama-Romney Presidential Debate
    </a>
   </p>
  </blockquote>
  <hr/>
  <p class="graytext">
   2008 Transcripts
  </p>
  <blockquote>
   <p>
    <a href="http://www.debates.org/index.php?page=2008-debate-transcript" title="September 26, 2008 De
bate Transcript">
     September 26, 2008: The First McCain-Obama Presidential Debate
    </a>
   </p>
   <p>
    <a href="http://www.debates.org/index.php?page=2008-debate-transcript-2" title="October 2, 2008 Deb
ate Transcript">
     October 2, 2008: The Biden-Palin Vice Presidential Debate
    </a>
   </p>
   <p>
    <a href="http://www.debates.org/index.php?page=october-7-2008-debate-transcrip" title="October 7, 2
008 Debate Transcript">
     October 7, 2008: The Second McCain-Obama Presidential Debate
    </a>
   </p>
   <p>
    <a href="http://www.debates.org/index.php?page=october-15-2008-debate-transcript" title="October 15
, 2008 Debate Transcript">
     October 15, 2008: The Third McCain-Obama Presidential Debate
    </a>
   </p>
  </blockquote>
  <hr/>
  <p class="graytext">
   2004 Transcripts
  </p>
  <blockquote>
   <p>
    <a href="http://www.debates.org/index.php?page=october-13-2004-debate-transcript" title="October 13
, 2004 Debate Transcript">
     October 13, 2004: The Third Bush-Kerry Presidential Debate
    </a>
   </p>
   <p>
    <a href="http://www.debates.org/index.php?page=october-8-2004-debate-transcript" title="October 8,
2004 Debate Transcript">
     October 8, 2004: The Second Bush-Kerry Presidential Debate
    </a>
   </p>
   <p>
    <a href="http://www.debates.org/index.php?page=october-5-2004-transcript" title="October 5, 2004 Tr
anscript">
     October 5, 2004: The Cheney-Edwards Vice Presidential Debate
    </a>
   </p>
   <p>
    <a href="http://www.debates.org/index.php?page=september-30-2004-debate-transcript" title="Septembe
r 30. 2004 Debate Transcript">
     September 30, 2004: The First Bush-Kerry Presidential Debate
    </a>
   </p>
  </blockquote>
  <hr/>
  <p class="graytext">
   2000 Transcripts
  </p>
  <blockquote>
   <p>
    <a href="http://www.debates.org/index.php?page=october-3-2000-transcript" title="October 3, 2000 Tr
anscript">
     October 3, 2000: The First Gore-Bush Presidential Debate
    </a>
   </p>
   <p>
    <a href="http://www.debates.org/index.php?page=october-5-2000-debate-transcript" title="October 5,
2000 Debate Transcript">
     October 5, 2000: The Lieberman-Cheney Vice Presidential Debate
    </a>
   </p>
   <p>
    <a href="http://www.debates.org/index.php?page=october-11-2000-debate-transcript" title="October 11
, 2000 Debate Transcript">
     October 11, 2000: The Second Gore-Bush Presidential Debate
    </a>
   </p>
   <p>
    <a href="http://www.debates.org/index.php?page=october-17-2000-debate-transcript" title="October 17
, 2000 Debate Transcript">
     October 17, 2000: The Third Gore-Bush Presidential Debate
    </a>
   </p>
   <p>
    <a href="http://www.debates.org/index.php?page=2000-debate-transcripts-translations" title="Debate
Transcript Translations">
     The 2000 Debate Transcripts: Transcripts of the debates translated into six languages
    </a>
   </p>
  </blockquote>
  <hr/>
  <p class="graytext">
   1996 Transcripts
  </p>
  <blockquote>
   <p>
    <a href="http://www.debates.org/index.php?page=october-6-1996-debate-transcript" title="October 6,
1996 Debate Transcript">
     October 6, 1996: The First Clinton-Dole Presidential Debate
    </a>
   </p>
   <p>
    <a href="http://www.debates.org/index.php?page=october-9-1996-debate-transcript" title="October 9,
1996 Debate Transcript">
     October 9, 1996: The Gore-Kemp Vice Presidential Debate
    </a>
   </p>
   <p>
    <a href="http://www.debates.org/index.php?page=october-16-1996-debate-transcript" title="October 16
, 1996 Debate Transcript">
     October 16, 1996: The Second Clinton-Dole Presidential Debate
    </a>
   </p>
  </blockquote>
  <hr/>
  <p class="graytext">
   1992 Transcripts
  </p>
  <blockquote>
   <p>
    October 11, 1992: The First Clinton-Bush-Perot Presidential Debate
   </p>
   <p>
    <a href="http://www.debates.org/index.php?page=october-11-1992-first-half-debate-transcript" title=
"October 11, 1992 First Half Debate Transcript">
     First half of Debate
    </a>
    <br/>
    <a href="http://www.debates.org/index.php?page=october-11-1992-second-half-debate-transcript" title
="October 11, 1992 Second Half Debate Transcript">
```

```
     Second half of Debate
    </a>
   </p>
   <p>
    <a href="http://www.debates.org/index.php?page=october-13-1992-debate-transcript" title="October 13
, 1992 Debate Transcript">
     October 13, 1992: The Gore-Quayle-Stockdale Vice Presidential Debate
    </a>
   </p>
   <p>
    October 15, 1992: The Second Clinton-Bush-Perot Presidential Debate
   </p>
   <p>
    <a href="http://www.debates.org/index.php?page=october-15-1992-first-half-debate-transcript" title=
"October 15, 1992 First Half Debate Transcript">
     First half of Debate
    </a>
    <br/>
    <a href="http://www.debates.org/index.php?page=october-15-1992-second-half-debate-transcript" title
="October 15, 1992 Second Half Debate Transcript">
     Second half of Debate
    </a>
   </p>
   <p>
    <a href="http://www.debates.org/index.php?page=october-19-1992-debate-transcript" title="October 19
, 1992 Debate Transcript">
     October 19, 1992: The Third Clinton-Bush-Perot Presidential Debate
    </a>
   </p>
  </blockquote>
  <hr/>
  <p class="graytext">
   1988 Transcripts
  </p>
  <blockquote>
   <p>
    <a href="http://www.debates.org/index.php?page=september-25-1988-debate-transcript" title="Septembe
r 25, 1988 Debate Transcript">
     September 25, 1988: The First Bush-Dukakis Presidential Debate
    </a>
   </p>
   <p>
    <a href="http://www.debates.org/index.php?page=october-5-1988-debate-transcripts" title="October 5,
1988 Debate Transcripts">
     October 5, 1988: The Bentsen-Quayle Vice Presidential Debate
    </a>
   </p>
   <p>
    <a href="http://www.debates.org/index.php?page=october-13-1988-debate-transcript" title="October 13
, 1988 Debate Transcript">
     October 13, 1988: The Second Bush-Dukakis Presidential Debate
    </a>
   </p>
  </blockquote>
  <hr/>
  <p class="graytext">
   1984 Transcripts
  </p>
  <blockquote>
   <p>
    <a href="http://www.debates.org/index.php?page=october-7-1984-debate-transcript" title="October 7,
1984 Debate Transcript">
     October 7, 1984: The First Reagan-Mondale Presidential Debate
    </a>
   </p>
   <p>
    <a href="http://www.debates.org/index.php?page=october-11-1984-debate-transcript" title="October 11
, 1984 Debate Transcript">
     October 11, 1984: The Bush-Ferraro Vice Presidential Debate
    </a>
   </p>
   <p>
    <a href="http://www.debates.org/index.php?page=october-21-1984-debate-transcript" title="October 21
, 1984 Debate Transcript">
     October 21, 1984: The Second Reagan-Mondale Presidential Debate
    </a>
   </p>
  </blockquote>
  <hr/>
  <p class="graytext">
   1980 Transcripts
  </p>
  <blockquote>
   <p>
    <a href="http://www.debates.org/index.php?page=september-21-1980-debate-transcript" title="Septembe
r 21, 1980 Debate Transcript">
     September 21, 1980: The Anderson-Reagan Presidential Debate
    </a>
   </p>
   <p>
    <a href="http://www.debates.org/index.php?page=october-28-1980-debate-transcript" title="October 28
, 1980 Debate Transcript">
     October 28, 1980: The Carter-Reagan Presidential Debate
    </a>
   </p>
  </blockquote>
  <hr/>
  <p class="graytext">
   1976 Transcripts
  </p>
  <blockquote>
   <dl>
    <dt>
     <a href="http://www.debates.org/index.php?page=september-23-1976-debate-transcript" title="Septemb
er 23, 1976 Debate Transcript">
      September 23, 1976: The First Carter-Ford Presidential Debate
     </a>
    </dt>
    <dt>
     <br/>
    </dt>
    <dt>
     <a href="http://www.debates.org/index.php?page=october-6-1976-debate-transcript" title="October 6,
1976 Debate Transcript">
      October 6, 1976: The Second Carter-Ford Presidential Debate
     </a>
    </dt>
    <dt>
     <br/>
    </dt>
    <dt>
     <a href="http://www.debates.org/index.php?page=october-22-1976-debate-transcript" title="October 2
2, 1976 Debate Transcript">
      October 22, 1976: The Third Carter-Ford Presidential Debate
     </a>
    </dt>
   </dl>
  </blockquote>
  <hr/>
  <p class="graytext">
   1960 Transcripts
  </p>
  <blockquote>
   <p>
    <a href="http://www.debates.org/index.php?page=september-26-1960-debate-transcript" title="Septembe
r 26, 1960 Debate Transcript">
     September 26, 1960: The First Kennedy-Nixon Presidential Debate
    </a>
   </p>
   <p>
    <a href="http://www.debates.org/index.php?page=october-7-1960-debate-transcript" title="October 7,
1960 Debate Transcript">
     October 7, 1960: The Second Kennedy-Nixon Presidential Debate
    </a>
   </p>
   <p>
    <a href="http://www.debates.org/index.php?page=october-13-1960-debate-transcript" title="October 13
, 1960 Debate Transcript">
     October 13, 1960: The Third Kennedy-Nixon Presidential Debate
    </a>
   </p>
   <p>
    <a href="http://www.debates.org/index.php?page=october-21-1960-debate-transcript" title="October 21
, 1960 Debate Transcript">
     October 21, 1960: The Fourth Kennedy-Nixon Presidential Debate
    </a>
   </p>
  </blockquote>
  <br/>
 </p>
</div>
```

In [62]:
```
# find all the links
link_result = list()
colname = list()
for p in Debates.find_all('a'):
    if 'The First' in p.text:
        link_result.append(p.get('href'))
        colname.append(p.text)
link_result
```

Out[62]:
```
['http://www.debates.org/index.php?page=october-3-2012-debate-transcript',
 'http://www.debates.org/index.php?page=2008-debate-transcript',
 'http://www.debates.org/index.php?page=september-30-2004-debate-transcript',
 'http://www.debates.org/index.php?page=october-3-2000-transcript',
 'http://www.debates.org/index.php?page=october-6-1996-debate-transcript',
 'http://www.debates.org/index.php?page=september-25-1988-debate-transcript',
 'http://www.debates.org/index.php?page=october-7-1984-debate-transcript',
 'http://www.debates.org/index.php?page=september-23-1976-debate-transcript',
 'http://www.debates.org/index.php?page=september-26-1960-debate-transcript']
```

In [63]:
```
pd.DataFrame(columns=colname)
```

Out[63]:

	October 3, 2012: The First Obama-Romney Presidential Debate	September 26, 2008: The First McCain-Obama Presidential Debate	September 30, 2004: The First Bush-Kerry Presidential Debate	October 3, 2000: The First Gore-Bush Presidential Debate	October 6, 1996: The First Clinton-Dole Presidential Debate	September 25, 1988: The First Bush-Dukakis Presidential Debate	October 7, 1984: The First Reagan-Mondale Presidential Debate	September 23, 1976: The First Carter-Ford Presidential Debate	September 26, 1960: The First Kennedy-Nixon Presidential Debate

In [64]:
```
# Count how long the transcript of the debate is
length_list = list()
for l in link_result:
    source = requests.get(l)
    soup = bs.BeautifulSoup(source.content, features='html.parser')
    text = soup.text.replace("\n","")
    length_list.append(len(text))
length_list
```

Out[64]:
```
[97403, 185201, 85500, 93835, 95866, 90273, 89463, 83516, 63716]
```

In [65]:
```
# Count how many times the word war was used in the different debates
war_count = list()
for l in link_result:
    s = requests.get(l)
    soup = bs.BeautifulSoup(s.content, features='html.parser')
    plain_text = soup.text
    plain_text = plain_text.replace("\n","").lower()
    found = re.findall(r'\bwar(?:es|s)?\b', plain_text)
    war_count.append(len(found))
war_count
```

Out[65]:
```
[5, 48, 64, 11, 15, 14, 3, 7, 3]
```

In [66]:
```
# scrape the most common used word in the debate, and write how many times it was used
most_common_w = list()
mcw_count=list()
for l in link_result:
    s = requests.get(l)
    soup = bs.BeautifulSoup(s.content, features='html.parser')
    plain_text = soup.text
    plain_text = plain_text.replace("\n","").lower()
    c = Counter(re.findall(r"[\w']+", plain_text.lower()))
    most_common_w.append((c.most_common()[0][0]))
    mcw_count.append((c.most_common()[0][1]))
```

In [67]:
```
most_common_w
```

Out[67]:
```
['the', 'the', 'the', 'the', 'the', 'the', 'the', 'the', 'the']
```

```
In [68]: mcw_count

Out[68]: [758, 1471, 858, 920, 877, 804, 866, 857, 779]
```

```
In [69]: output=pd.DataFrame([length_list,war_count,most_common_w,mcw_count],columns=colname,
                             index=['Debates char length','War Count','Most common w','Most common w count'])
         output

Out[69]:
```

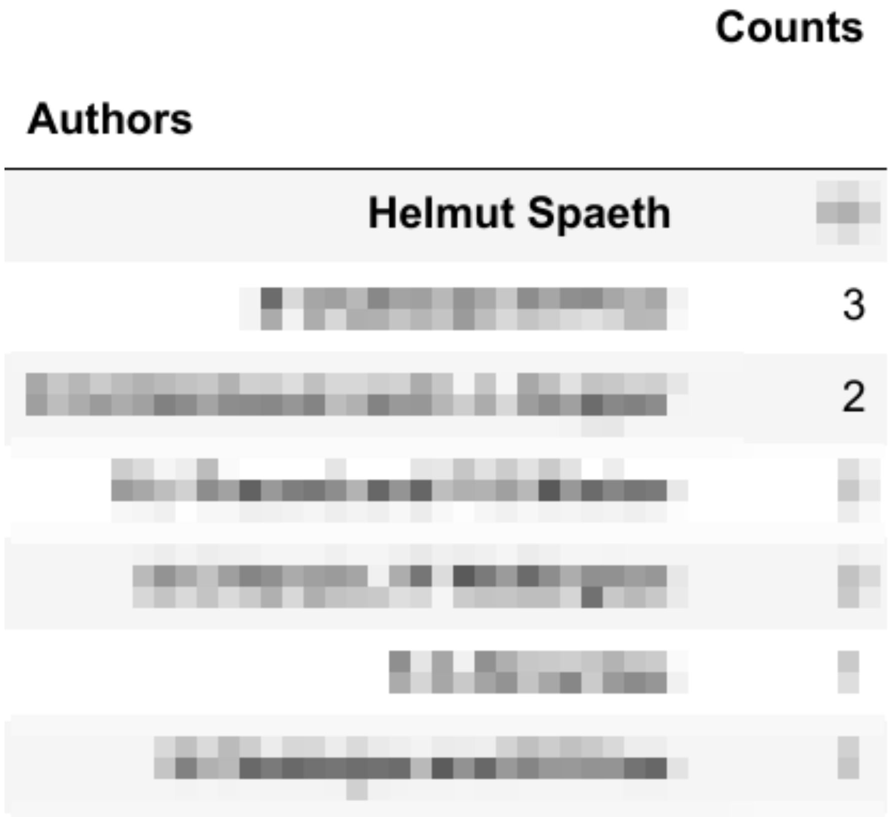
	October 3, 2012: The First Obama-Romney Presidential Debate	September 26, 2008: The First McCain-Obama Presidential Debate	September 30, 2004: The First Bush-Kerry Presidential Debate	October 3, 2000: The First Gore-Bush Presidential Debate	October 6, 1996: The First Clinton-Dole Presidential Debate	September 25, 1988: The First Bush-Dukakis Presidential Debate	October 7, 1984: The First Reagan-Mondale Presidential Debate	September 23, 1976: The First Carter-Ford Presidential Debate	September 26, 1960: The First Kennedy-Nixon Presidential Debate
Debates char length	97403	185201	85500	93835	95866	90273	89463	83516	63716
War Count	5	48	64	11	15	14	3	7	3
Most common w	the	the	the	the	the	the	the	the	the
Most common w count	758	1471	858	920	877	804	866	857	779

2. Download and read in specific line from many data sets

Scrape the first 27 data sets from this URL http://people.sc.fsu.edu/~jburkardt/datasets/regression/ (http://people.sc.fsu.edu/~jburkardt/datasets/regression/) (i.e.`x01.txt` - `x27.txt`). Then, save the 5th line in each data set, this should be the name of the data set author (get rid of the # symbol, the white spaces and the comma at the end).

Count how many times (with a Python function) each author is the reference for one of the 27 data sets. Showcase your results, sorted, with the most common author name first and how many times he appeared in data sets. Use a Pandas DataFrame to show your results, see example. Print your final output result.

Example output of the answer for Question 2:



	Counts
Authors	
Helmut Spaeth	
	3
	2

```
In [70]: source2 = requests.get("http://people.sc.fsu.edu/~jburkardt/datasets/regression/")
         soup2 = bs.BeautifulSoup(source2.content, features='html.parser')
         print(soup2.body)

         <body>
```

```
<h1>Index of /~jburkardt/datasets/regression</h1>
<table><tr><th><img alt="[ICO]" src="/icons/blank.gif"/></th><th><a href="?C=N;O=D">Name</a></th><th><a
href="?C=M;O=A">Last modified</a></th><th><a href="?C=S;O=A">Size</a></th><th><a href="?C=D;O=A">Descri
ption</a></th></tr><tr><th colspan="5"><hr/></th></tr>
<tr><td valign="top"><img alt="[DIR]" src="/icons/back.gif"/></td><td><a href="/~jburkardt/datasets/">P
arent Directory</a></td><td> </td><td align="right">  - </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="regression.html">regress
ion.html</a></td><td align="right">11-Mar-2015 08:19  </td><td align="right"> 20K</td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x01.txt">x01.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">2.0K</td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x02.txt">x02.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">1.0K</td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x03.txt">x03.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">1.2K</td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x04.txt">x04.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">1.7K</td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x05.txt">x05.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">1.9K</td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x06.txt">x06.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">1.7K</td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x07.txt">x07.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">2.0K</td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x08.txt">x08.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">1.4K</td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x09.txt">x09.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">1.2K</td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x10.txt">x10.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">1.2K</td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x11.txt">x11.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">2.9K</td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x12.txt">x12.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">1.4K</td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x13.txt">x13.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">1.5K</td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x14.txt">x14.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">2.0K</td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x15.txt">x15.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">2.7K</td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x16.txt">x16.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">2.9K</td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x17.txt">x17.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">3.9K</td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x18.txt">x18.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">4.2K</td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x19.txt">x19.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">1.7K</td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x20.txt">x20.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">2.4K</td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x21.txt">x21.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">1.8K</td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x22.txt">x22.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">1.9K</td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x23.txt">x23.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">1.7K</td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x24.txt">x24.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">1.8K</td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x25.txt">x25.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">2.1K</td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x26.txt">x26.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">3.4K</td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x27.txt">x27.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">3.5K</td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x28.txt">x28.txt</a></td
><td align="right">08-Aug-2016 07:48  </td><td align="right">7.8K</td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x29.txt">x29.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">701 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x30.txt">x30.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">741 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x31.txt">x31.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">1.0K</td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x32.txt">x32.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">902 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x33.txt">x33.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">3.7K</td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x34.txt">x34.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">1.3K</td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x35.txt">x35.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">1.5K</td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x36.txt">x36.txt</a></td
```

```
><td align="right">03-Oct-2011 16:52  </td><td align="right">1.0K</td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x37.txt">x37.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">1.1K</td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x38.txt">x38.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">1.1K</td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x39.txt">x39.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">875 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x40.txt">x40.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">915 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x41.txt">x41.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">928 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x42.txt">x42.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">1.7K</td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x43.txt">x43.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">461 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x43_01.txt">x43_01.txt</
a></td><td align="right">03-Oct-2011 16:52  </td><td align="right">317 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x43_02.txt">x43_02.txt</
a></td><td align="right">03-Oct-2011 16:52  </td><td align="right">334 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x43_03.txt">x43_03.txt</
a></td><td align="right">03-Oct-2011 16:52  </td><td align="right">308 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x44.txt">x44.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">464 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x44_01.txt">x44_01.txt</
a></td><td align="right">03-Oct-2011 16:52  </td><td align="right">317 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x44_02.txt">x44_02.txt</
a></td><td align="right">03-Oct-2011 16:52  </td><td align="right">334 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x44_03.txt">x44_03.txt</
a></td><td align="right">03-Oct-2011 16:52  </td><td align="right">308 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x45.txt">x45.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">473 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x45_01.txt">x45_01.txt</
a></td><td align="right">03-Oct-2011 16:52  </td><td align="right">317 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x45_02.txt">x45_02.txt</
a></td><td align="right">03-Oct-2011 16:52  </td><td align="right">334 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x45_03.txt">x45_03.txt</
a></td><td align="right">03-Oct-2011 16:52  </td><td align="right">308 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x46.txt">x46.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">270 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x47.txt">x47.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">863 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x47_01.txt">x47_01.txt</
a></td><td align="right">03-Oct-2011 16:52  </td><td align="right">316 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x47_02.txt">x47_02.txt</
a></td><td align="right">03-Oct-2011 16:52  </td><td align="right">273 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x47_03.txt">x47_03.txt</
a></td><td align="right">03-Oct-2011 16:52  </td><td align="right">289 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x48.txt">x48.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">855 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x48_01.txt">x48_01.txt</
a></td><td align="right">03-Oct-2011 16:52  </td><td align="right">316 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x48_02.txt">x48_02.txt</
a></td><td align="right">03-Oct-2011 16:52  </td><td align="right">273 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x48_03.txt">x48_03.txt</
a></td><td align="right">03-Oct-2011 16:52  </td><td align="right">285 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x49.txt">x49.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">805 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x49_01.txt">x49_01.txt</
a></td><td align="right">03-Oct-2011 16:52  </td><td align="right">316 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x49_02.txt">x49_02.txt</
a></td><td align="right">03-Oct-2011 16:52  </td><td align="right">285 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x49_03.txt">x49_03.txt</
a></td><td align="right">03-Oct-2011 16:52  </td><td align="right">266 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x50.txt">x50.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">803 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x50_01.txt">x50_01.txt</
a></td><td align="right">03-Oct-2011 16:52  </td><td align="right">316 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x50_02.txt">x50_02.txt</
a></td><td align="right">03-Oct-2011 16:52  </td><td align="right">285 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x50_03.txt">x50_03.txt</
a></td><td align="right">03-Oct-2011 16:52  </td><td align="right">266 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x51.txt">x51.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">803 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x51_01.txt">x51_01.txt</
a></td><td align="right">03-Oct-2011 16:52  </td><td align="right">316 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x51_02.txt">x51_02.txt</
a></td><td align="right">03-Oct-2011 16:52  </td><td align="right">285 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x51_03.txt">x51_03.txt</
a></td><td align="right">03-Oct-2011 16:52  </td><td align="right">287 </td></tr>
```

```
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x52.txt">x52.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">803 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x52_01.txt">x52_01.txt</
a></td><td align="right">03-Oct-2011 16:52  </td><td align="right">316 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x52_02.txt">x52_02.txt</
a></td><td align="right">03-Oct-2011 16:52  </td><td align="right">285 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x52_03.txt">x52_03.txt</
a></td><td align="right">03-Oct-2011 16:52  </td><td align="right">287 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x53.txt">x53.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">805 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x53_01.txt">x53_01.txt</
a></td><td align="right">03-Oct-2011 16:52  </td><td align="right">316 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x53_02.txt">x53_02.txt</
a></td><td align="right">03-Oct-2011 16:52  </td><td align="right">285 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x53_03.txt">x53_03.txt</
a></td><td align="right">03-Oct-2011 16:52  </td><td align="right">287 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x54.txt">x54.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">889 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x54_01.txt">x54_01.txt</
a></td><td align="right">03-Oct-2011 16:52  </td><td align="right">471 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x54_02.txt">x54_02.txt</
a></td><td align="right">03-Oct-2011 16:52  </td><td align="right">347 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x54_03.txt">x54_03.txt</
a></td><td align="right">03-Oct-2011 16:52  </td><td align="right">318 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x55.txt">x55.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">950 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x55_01.txt">x55_01.txt</
a></td><td align="right">03-Oct-2011 16:52  </td><td align="right">535 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x55_02.txt">x55_02.txt</
a></td><td align="right">03-Oct-2011 16:52  </td><td align="right">287 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x55_03.txt">x55_03.txt</
a></td><td align="right">03-Oct-2011 16:52  </td><td align="right">410 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x56.txt">x56.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">815 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x56_01.txt">x56_01.txt</
a></td><td align="right">03-Oct-2011 16:52  </td><td align="right">451 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x56_02.txt">x56_02.txt</
a></td><td align="right">03-Oct-2011 16:52  </td><td align="right">281 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x56_03.txt">x56_03.txt</
a></td><td align="right">03-Oct-2011 16:52  </td><td align="right">379 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x57.txt">x57.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">815 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x57_01.txt">x57_01.txt</
a></td><td align="right">03-Oct-2011 16:52  </td><td align="right">469 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x57_02.txt">x57_02.txt</
a></td><td align="right">03-Oct-2011 16:52  </td><td align="right">281 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x57_03.txt">x57_03.txt</
a></td><td align="right">03-Oct-2011 16:52  </td><td align="right">379 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x58.txt">x58.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">852 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x58_01.txt">x58_01.txt</
a></td><td align="right">03-Oct-2011 16:52  </td><td align="right">331 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x58_03.txt">x58_03.txt</
a></td><td align="right">03-Oct-2011 16:52  </td><td align="right">298 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x59.txt">x59.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">860 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x59_01.txt">x59_01.txt</
a></td><td align="right">03-Oct-2011 16:52  </td><td align="right">322 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x59_02.txt">x59_02.txt</
a></td><td align="right">03-Oct-2011 16:52  </td><td align="right">283 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x59_03.txt">x59_03.txt</
a></td><td align="right">03-Oct-2011 16:52  </td><td align="right">291 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x60.txt">x60.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">648 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x61.txt">x61.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">1.0K</td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x61_01.txt">x61_01.txt</
a></td><td align="right">03-Oct-2011 16:52  </td><td align="right">534 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x61_02.txt">x61_02.txt</
a></td><td align="right">03-Oct-2011 16:52  </td><td align="right">418 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x61_03.txt">x61_03.txt</
a></td><td align="right">03-Oct-2011 16:52  </td><td align="right">389 </td></tr>
<tr><td valign="top"><img alt="[TXT]" src="/icons/text.gif"/></td><td><a href="x62.txt">x62.txt</a></td
><td align="right">03-Oct-2011 16:52  </td><td align="right">743 </td></tr>
<tr><th colspan="5"><hr/></th></tr>
</table>
</body>
```

In [71]:

```python
# Scrape the first 27 data sets
url_result=list()
for p in soup2.find_all('a'):
    if'.txt'in p.text:
        url_result.append('http://people.sc.fsu.edu/~jburkardt/datasets/regression/'+ p.get('href'))
url_result = url_result[0:27]
url_result
```

Out[71]:

```
['http://people.sc.fsu.edu/~jburkardt/datasets/regression/x01.txt',
 'http://people.sc.fsu.edu/~jburkardt/datasets/regression/x02.txt',
 'http://people.sc.fsu.edu/~jburkardt/datasets/regression/x03.txt',
 'http://people.sc.fsu.edu/~jburkardt/datasets/regression/x04.txt',
 'http://people.sc.fsu.edu/~jburkardt/datasets/regression/x05.txt',
 'http://people.sc.fsu.edu/~jburkardt/datasets/regression/x06.txt',
 'http://people.sc.fsu.edu/~jburkardt/datasets/regression/x07.txt',
 'http://people.sc.fsu.edu/~jburkardt/datasets/regression/x08.txt',
 'http://people.sc.fsu.edu/~jburkardt/datasets/regression/x09.txt',
 'http://people.sc.fsu.edu/~jburkardt/datasets/regression/x10.txt',
 'http://people.sc.fsu.edu/~jburkardt/datasets/regression/x11.txt',
 'http://people.sc.fsu.edu/~jburkardt/datasets/regression/x12.txt',
 'http://people.sc.fsu.edu/~jburkardt/datasets/regression/x13.txt',
 'http://people.sc.fsu.edu/~jburkardt/datasets/regression/x14.txt',
 'http://people.sc.fsu.edu/~jburkardt/datasets/regression/x15.txt',
 'http://people.sc.fsu.edu/~jburkardt/datasets/regression/x16.txt',
 'http://people.sc.fsu.edu/~jburkardt/datasets/regression/x17.txt',
 'http://people.sc.fsu.edu/~jburkardt/datasets/regression/x18.txt',
 'http://people.sc.fsu.edu/~jburkardt/datasets/regression/x19.txt',
 'http://people.sc.fsu.edu/~jburkardt/datasets/regression/x20.txt',
 'http://people.sc.fsu.edu/~jburkardt/datasets/regression/x21.txt',
 'http://people.sc.fsu.edu/~jburkardt/datasets/regression/x22.txt',
 'http://people.sc.fsu.edu/~jburkardt/datasets/regression/x23.txt',
 'http://people.sc.fsu.edu/~jburkardt/datasets/regression/x24.txt',
 'http://people.sc.fsu.edu/~jburkardt/datasets/regression/x25.txt',
 'http://people.sc.fsu.edu/~jburkardt/datasets/regression/x26.txt',
 'http://people.sc.fsu.edu/~jburkardt/datasets/regression/x27.txt']
```

In [72]:

```python
# save the 5th line in each data set (author name)
author_list=list()

for l in url_result:
    source = requests.get(l)
    plain_text = source.text
    line = re.sub('[!@#$,]','', plain_text.split('\n')[4])
    author_list.append(line.strip())
author_list
```

Out[72]:

```
['Helmut Spaeth',
 'Helmut Spaeth',
 'Helmut Spaeth',
 'Helmut Spaeth',
 'Helmut Spaeth',
 'R J Freund and P D Minton',
 'D G Kleinbaum and L L Kupper',
 'Helmut Spaeth',
 'D G Kleinbaum and L L Kupper',
 'K A Brownlee',
 'Helmut Spaeth',
 'Helmut Spaeth',
 'S Chatterjee and B Price',
 'Helmut Spaeth',
 'Helmut Spaeth',
 'Helmut Spaeth',
 'Helmut Spaeth',
 'Helmut Spaeth',
 'R J Freund and P D Minton',
 'Helmut Spaeth',
 'Helmut Spaeth',
 'Helmut Spaeth',
 'S Chatterjee B Price',
 'S Chatterjee B Price',
 'S Chatterjee B Price',
 'S C Narula J F Wellington',
 'S C Narula J F Wellington']
```

```
In [73]: author_list= [y  for x in author_list for y in x.split(' and ')]
         author_list
```

```
Out[73]: ['Helmut Spaeth',
          'Helmut Spaeth',
          'Helmut Spaeth',
          'Helmut Spaeth',
          'Helmut Spaeth',
          'R J Freund',
          'P D Minton',
          'D G Kleinbaum',
          'L L Kupper',
          'Helmut Spaeth',
          'D G Kleinbaum',
          'L L Kupper',
          'K A Brownlee',
          'Helmut Spaeth',
          'Helmut Spaeth',
          'S Chatterjee',
          'B Price',
          'Helmut Spaeth',
          'Helmut Spaeth',
          'Helmut Spaeth',
          'Helmut Spaeth',
          'Helmut Spaeth',
          'R J Freund',
          'P D Minton',
          'Helmut Spaeth',
          'Helmut Spaeth',
          'Helmut Spaeth',
          'S Chatterjee B Price',
          'S Chatterjee B Price',
          'S Chatterjee B Price',
          'S C Narula J F Wellington',
          'S C Narula J F Wellington']
```

```
In [74]: author_str = ' '.join(author_list)
         author_str
```

```
Out[74]: 'Helmut Spaeth Helmut Spaeth Helmut Spaeth Helmut Spaeth Helmut Spaeth R J Freund P D Minton D G Kleinb
         aum L L Kupper Helmut Spaeth D G Kleinbaum L L Kupper K A Brownlee Helmut Spaeth Helmut Spaeth S Chatte
         rjee B Price Helmut Spaeth Helmut Spaeth Helmut Spaeth Helmut Spaeth Helmut Spaeth R J Freund P D Minto
         n Helmut Spaeth Helmut Spaeth Helmut Spaeth S Chatterjee B Price S Chatterjee B Price S Chatterjee B Pr
         ice S C Narula J F Wellington S C Narula J F Wellington'
```

```
In [75]: count_list =list()
         Author = ['Helmut Spaeth','R J Freund','P D Minton','D G Kleinbaum','B Price', 'L L Kupper','K A Brownle
         e','S Chatterjee','S C Narula J','F Wellington']
         for name in Author:
             count_list.append(author_str.count(name))
         count_list
```

```
Out[75]: [16, 2, 2, 2, 4, 2, 1, 4, 2, 2]
```

```
In [76]: a_df = pd.DataFrame({'Author': Author, 'Count': count_list})
         a_df
```

Out[76]:

	Author	Count
0	Helmut Spaeth	16
1	R J Freund	2
2	P D Minton	2
3	D G Kleinbaum	2
4	B Price	4
5	L L Kupper	2
6	K A Brownlee	1
7	S Chatterjee	4
8	S C Narula J	2
9	F Wellington	2

```
In [77]: a_df.sort_values('Count',ascending=False)
```

Out[77]:

	Author	Count
0	Helmut Spaeth	16
4	B Price	4
7	S Chatterjee	4
1	R J Freund	2
2	P D Minton	2
3	D G Kleinbaum	2
5	L L Kupper	2
8	S C Narula J	2
9	F Wellington	2
6	K A Brownlee	1