Python for Social Science
Ling 596: Special Topics
Spring, 2014: No prerequisites
gawron@mail.sdsu.edu
TuTh 1400-1515 PSFA 413

Jean Mark Gawron
San Diego State University, Department of Linguistics

2013-28-10
Overview

1. Introduction
2. Visualization
3. Conclusion
4. Bibliography
Who am I?

1. A professor in the Department of Linguistics specializing in **Computational linguistics**
3. I have a lot of experience in introducing students without a lot of computational background to computational ideas
What am I plugging?

Python for Social Science
Ling 596: Special Topics
Spring, 2014: No prerequisites
TuTh 1400-1515 PSFA 413
Need

Big data and social science

Social scientists find themselves facing exponentially larger data sets without suitable tools to deal with them.
Where do you fit?

UC Berkeley Data Science Master’s
What is Python?

A programming language

1. Active and growing community of (data) scientists using it
2. Easy to learn
3. Easily constructed **scripts**: programs that construct pipelines combining the functionality of other programs
4. Provides a formidable array of data collection, data manipulation, and data analysis tools
Class goals

Who it’s for

1. Graduate students and upper division undergraduates
2. Students with no knowledge of programming who want to get in on the data goldmine of the Age of Information
3. Students who have data that they need to drill into to reshape it or to extract specific kinds of information.
4. Students open to expanding their computational skills
Class prereqs

1. Some knowledge of what counts as interesting data in your particular discipline, and some experience working with it.
2. An interest in exploring some of the data opportunities provided by government websites, social networks, blogs, and the marketplace of ideas that is the Internet.
<table>
<thead>
<tr>
<th>Topics</th>
<th>1. Python Basics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text Data</td>
<td>2. Searching for patterns in text and web data (regular expressions)</td>
</tr>
<tr>
<td></td>
<td>3. Extracting information from big data sources (Government data)</td>
</tr>
<tr>
<td>Analysis/visualization</td>
<td>4. Constructing social networks from data (visualizing social groups)</td>
</tr>
<tr>
<td></td>
<td>5. Connecting to your stat package (Python data frames)</td>
</tr>
<tr>
<td></td>
<td>6. Visualizing similarity relations</td>
</tr>
<tr>
<td></td>
<td>7. Visualizing quantitative relationships on maps</td>
</tr>
</tbody>
</table>
Data sources

1. PUMS (US Census)
2. Social Security Administration
3. Enron email data
4. Geocoding servers (Google) and geocoding DBs
5. USDA Food Database
6. Twitter
7. RSS news feeds
Outline

1. Introduction
2. Visualization
3. Conclusion
4. Bibliography
Visualization through similarity relations

Voting patterns

Coordinate 1

Coordinate 2
Social networks

Zachary (1977)
Enron email network

Shetty and Adibi (2005)

Figure 6: Enron Network
# Mapping trends I

Federal election contributions

<table>
<thead>
<tr>
<th>State</th>
<th>Obama, Barack</th>
<th>Romney, Mitt</th>
</tr>
</thead>
<tbody>
<tr>
<td>AK</td>
<td>275353.15</td>
<td>86204.24</td>
</tr>
<tr>
<td>AL</td>
<td>537835.48</td>
<td>504882.08</td>
</tr>
<tr>
<td>AR</td>
<td>324802.28</td>
<td>105351.50</td>
</tr>
<tr>
<td>AZ</td>
<td>1484241.84</td>
<td>1850484.23</td>
</tr>
<tr>
<td>CA</td>
<td>23370680.84</td>
<td>10908232.46</td>
</tr>
</tbody>
</table>
Mapping trends II
1. Python for Social Scientists is a new course addressing an interdisciplinary need
2. Graduate students and upper division undergraduates welcome
3. Python is freely available and may already be installed on your home computer (Macs, Linux)
4. A full course outline
5. Email gawron@mail.sdsu.edu with questions
Shetty, Jitesh, and Jafar Adibi. 2005.  
Discovering important nodes through graph entropy the case of enron email database.  
In Proceedings of the 3rd international workshop on Link discovery, 74–81. ACM.

An information flow model for conflict and fission in small groups.  
*Journal of Anthropological Research* 33:452–473.