Python for Social Science

Ling 572: Python Scripting

Fall, 2020: No prerequisites

gawron@sdsu.edu

TuTh 1600-1840

Jean Mark Gawron

San Diego State University, Department of Linguistics

2020-25-08

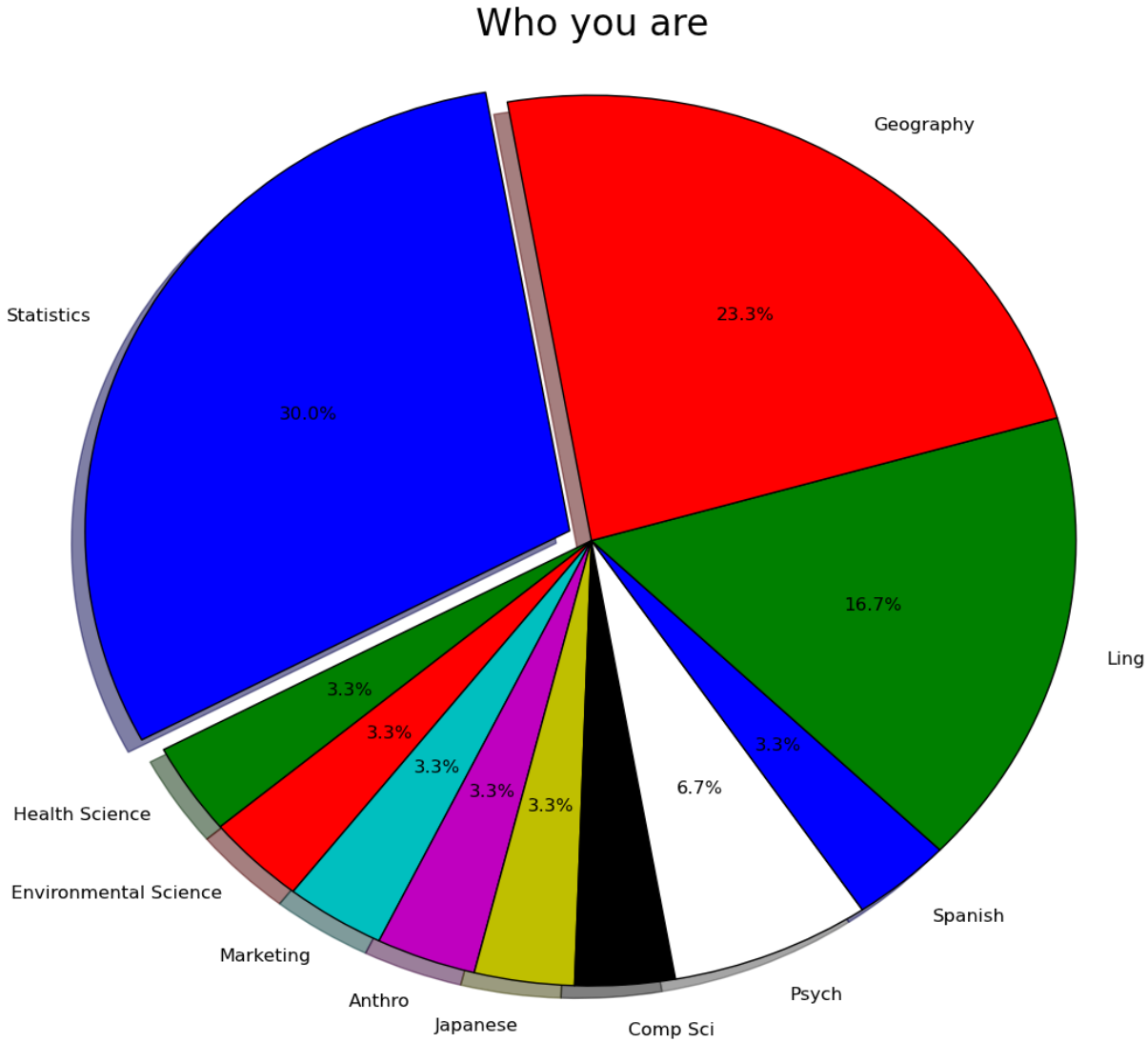
Overview

Outline

Who am I?

1. A professor in the Department of Linguistics specializing in **Computational linguistics**
2. Machine translation, Speech recognition, Text classification, Topic identification
3. I have a lot of experience in introducing students without a lot of computational background to computational ideas

The code

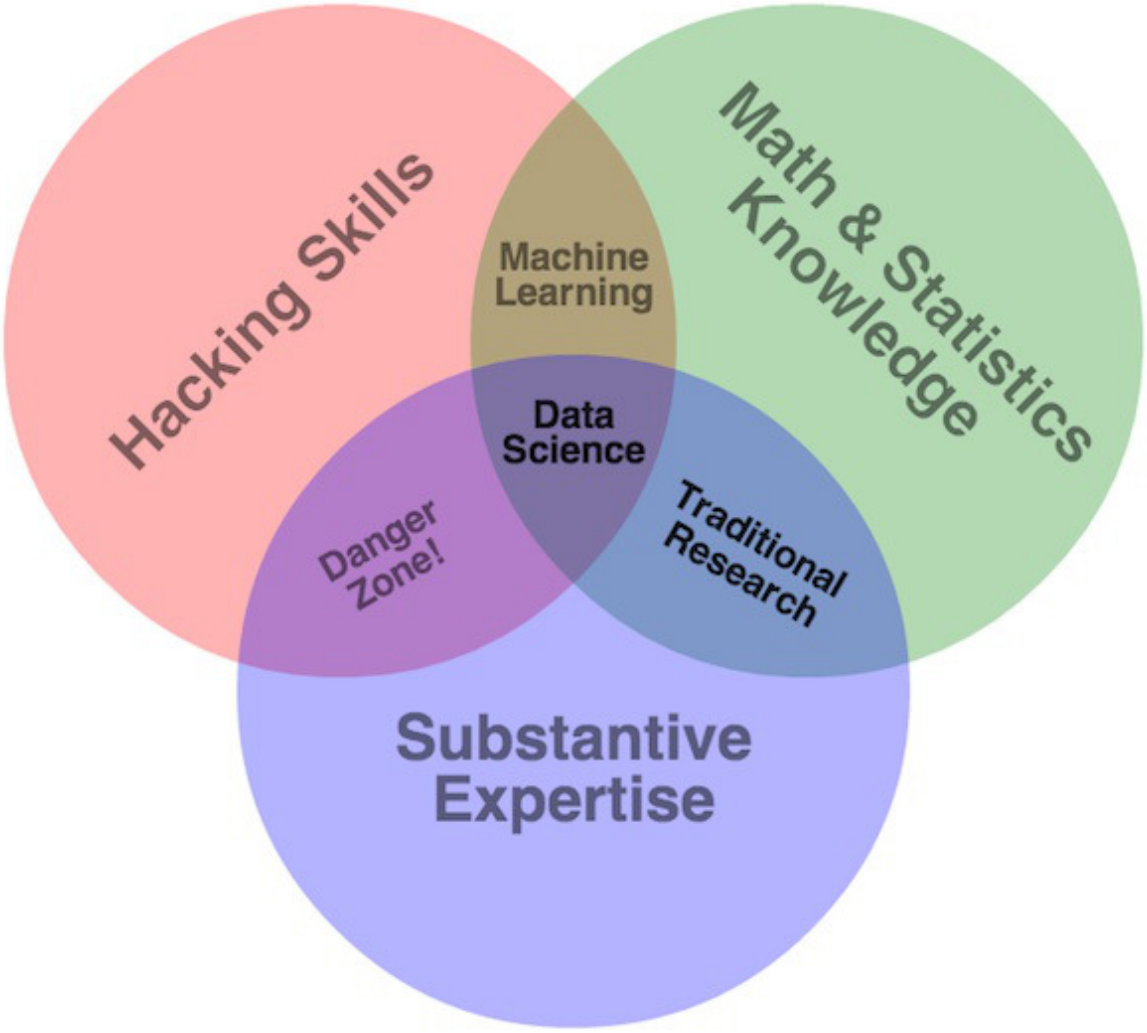

Who you are
Geography
23.3%
Statistics
30.0%
16.7%
Ling
3.3%
3.3%
3.3%
3.3%
3.3%
6.7%
3.3%
Health Science
Environmental Science
Marketing
Anthro
Japanese
Comp Sci
Psych
Spanish

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?



Drew Conways's Data Science Venn Diagram

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**

Outline

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.

Topics

	1. Python Basics
Numpy & Pandas	2. Arrays, array operations 3. Pandas data frames 4. Aggregation, analysis wuth dataframes
Machine Learning	3. General machine learning perspective (analytical tool) 3. Classifying texts
Visualization	5. Constructing social networks from data (visualizing social groups) 6. Python plotting tools 7. Visualizing quantitative relationships on maps 8. Visualizing similarity relations