

# Intro to Python and Web Scraping

**Soubhik Barari**

We the People / Hack for Democracy

MIT GOV/LAB

April 22, 2017

## Intro to Python and Web Scraping

Repository for setup instructions, notebooks, and slides:

<https://github.com/Hack4Democracy/Python-Workshop>  
[\(https://github.com/Hack4Democracy/Python-Workshop\)](https://github.com/Hack4Democracy/Python-Workshop)

## Overview

**Goal:** Teach you how to use Python for common purpose web scraping tasks!

**Subgoal:** Teach you coding skills to crush any hackathon.

### **Skills:**

1. Python Fundamentals
2. Web Scraping
  - HTML / XML
  - Working with Twitter data

## Python Fundamentals : Introduction

## Motivations

- How do I perform text analysis on a collection of documents?
- *How do I build an automated web scraping pipeline?*
- How do I write my own machine learning routines?

... How can I quickly develop and iterate on my computational ideas?



## Python Fundamentals : Introduction

### Why use Python?

- Highly general purpose (batteries included)
- Easily customizable (write anything you want)
- Functional *and* object-oriented (write any way you want)

## Python Fundamentals : Best Practices

### Installation recommendations

- For the best community support and third-party accessibility... install **Version 2.7**.
- For an off-the-shelf, all-inclusive data science environment... install **Anaconda**.

## Python Fundamentals : Best Practices

### Usage recommendations

- I want to write some quick, dirty commands ... use a **Python interpreter**.
- I want to create a beautiful, interactive document ... use a **Jupyter notebook**.
- I want to develop some *serious* software ... use an **IDE**:
  - "I want to build my own IDE!" ----- **Sublime Text**
  - "I'm lazy." ----- **Spyder**

# Python Fundamentals : Best Practices

# Interpreter

```
coulomm@laptop:~$ python
Python 2.7.12 (default, Jun 29 2016, 11:48:22)
[GCC 4.2.1 Compatible Apple LLVM 7.3.0 (Clang-703.0.31)] on darwin
Type "help", "copyright", "credits" or "license" for more information.
>>> import datetime as dt
>>> print dt.datetime.today()
2017-04-20 12:57:33.436108
>>> 
```

# Python Fundamentals : Best Practices

## Notebook

The screenshot shows a Jupyter Notebook interface with the following details:

- Title Bar:** The title bar says "jupyter python\_basic Last Checkpoint: Last Tuesday at 5:12 PM (unsaved changes)".
- Toolbar:** Standard Jupyter toolbar icons for File, Edit, View, Insert, Cell, Kernel, Help, and Cell Toolbar.
- Header:** Python 2.0
- Section Headers:** "Python Fundamentals" and "Overview".
- Text:** Welcome in this Jupyter notebook, we are going to go over some fundamentals of the Python programming language.
- Text:** When we're learning a new language, we usually start by writing a one-line program that prints the message "Hello world!". This is a simple program that shows whether your computer is properly set up to run Python programs.
- Code Cell:** In [1]:

```
print "Hello world!"  
print "This is so fun! Real time output! La la la la!"
```

Output:  
Hello world!  
This is so fun! Real time output! La la la la
- Text:** If the correct output "Hello world!" was displayed ... congrats, you've just ran your first Python program.
- Section Headers:** "Basic types".
- Text:** First we're going to go over the most fundamental building blocks of Python which are the basic unit data types.
- Text:** As you would expect, Python has a data type for integers (`int`), strings (`str`), booleans (`bool`), and for floating point numbers/decimals (`float`). If you're familiar with R, these types pretty much directly translate with minor syntactical differences.
- Code Cell:** In [1]:

```
# You can write comments with a single '#'  
  
a = 5  
b = "Hello world"  
c = 42.0  
d = ...  
e = True
```
- Text:** To print the value of an object, simply use the `print` function:
- Code Cell:** In [1]:

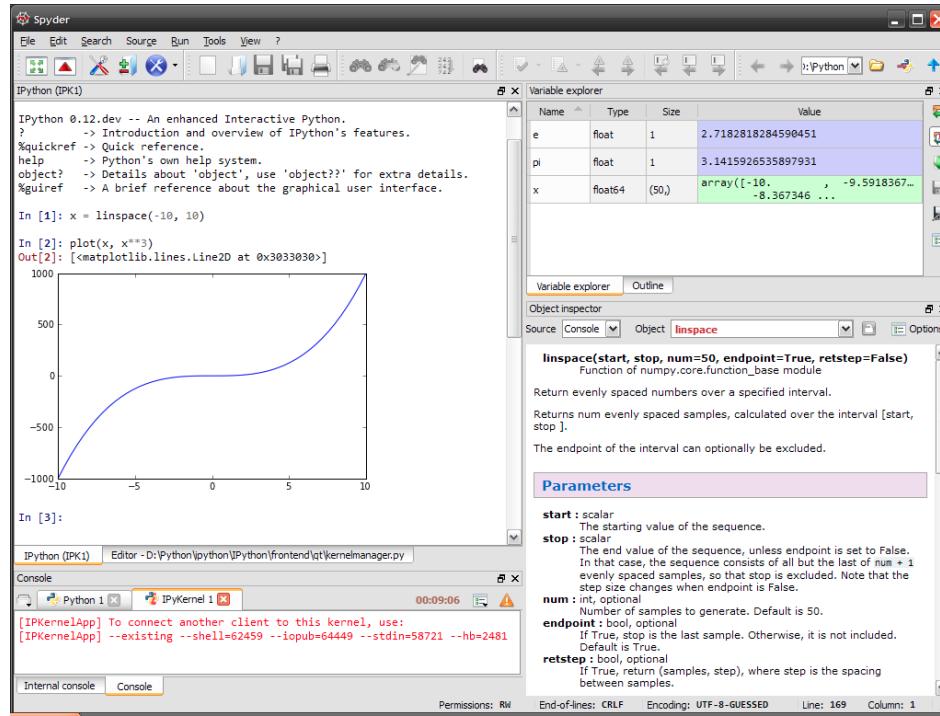
```
print a  
print b  
print c  
print e
```
- Text:** To print the type of an object, simple use the `type` function:
- Code Cell:** In [1]:

```
print type(a)  
print type(b)  
print type(c)  
print type(e)
```

Output:  
<type 'int'>  
<type 'str'>

# Python Fundamentals : Best Practices

## Spyder



# Python Fundamentals : Best Practices

## Package recommendations

- **pandas** : basic R-style data structures
- **numpy**, **statsmodels** : numerical / statistical computing
- **mechanize**, **beautifulsoup4**, **pdfminer** : web scraping
- **scikit-learn** : machine learning
- **networkx**, **graphx** : network analysis
- **bokeh**, **seaborn** : data visualization

# Python Fundamentals : Best Practices

The Zen of Python:

**Beautiful** is better than ugly

**Explicit** is better than implicit

**Simple** is better than complex

**Complex** is better than complicated

**Readability** counts

## Python Fundamentals : Language

### General syntax

```
In [1]: # This is a comment!

import math
import sys
import datetime as dt

def f(x): y = x*2; return y

# Don't need to declare types
x = "We're going to win bigly "
y = 41

if "bigly" in x or y == 42:
    # Indentation matters!
    x = x + ", I mean really bigly"
    y = y + 1

print x
print y
print math.pi
print dt.datetime.today()
```

```
We're going to win bigly , I mean really bigly
42
3.14159265359
2017-04-21 22:52:01.616947
```

## Python Fundamentals : Language

## Pandas

"Like SQL and R but in Python." *A must-learn for any data scientist!*

```
In [ ]: import pandas as pd

# Read in data
dataframe = pd.read_csv("myDataset.csv")

# Manipulate
men = dataframe[dataframe["sex"] == "male"]
women = dataframe[dataframe["sex"] == "female"]

# Output data
men.to_csv("myMen.csv")
women.to_csv("myWomen.csv")
```

## Python Fundamentals : Language

(see notebook)

## Python Fundamentals

### Tips

- Learn Python in 10 minutes (i.e. hackathon pace): <https://www.stavros.io/tutorials/python/> (<https://www.stavros.io/tutorials/python/>).
- Be comfortable with messy code.
- The best way to learn *any* new programming language:
  - 1.) Set up like everyone sets up and runs that language.
  - 2.) Learn the universal traits of that language (syntax, etc.).
  - 3.) Learn the idiosyncrasies of that language (e.g. functional programming).
  - 4.) Pick a project you're very excited about.
  - 5.) *Try to swim.*

## Web Scraping

### Web Scraping : Site structures

### Web Scraping : Site structures

HTML - "a block-based display language."

Display	Source																																																								
<p>MIT POLITICAL SCIENCE</p> <p>ABOUT   UNDERGRADUATE   GRADUATE   PEOPLE   RESEARCH   WORKING PAPERS   NEWS &amp; EVENTS</p> <p>Faculty   Affiliates   Staff   Visitors   Graduate Students</p> <p>PEOPLE   FACULTY</p> <table border="1"> <thead> <tr> <th>Name</th> <th>office</th> <th>phone</th> <th>email</th> </tr> </thead> <tbody> <tr><td>Regina Bateson</td><td>E53-451</td><td>324-7336</td><td>bateson@mit.edu</td></tr> <tr><td>Suzanne Berger</td><td>E53-451</td><td>253-6640</td><td>szberger@mit.edu</td></tr> <tr><td>Adam Berinsky</td><td>E53-457</td><td>253-8190</td><td>berinsky@mit.edu</td></tr> <tr><td>Donald Blackmer - Emeritus</td><td>E53-366</td><td>253-3145</td><td>blackmer@mit.edu</td></tr> <tr><td>Andrea Campbell - Department Head</td><td>E53-473</td><td>452-2295</td><td>acampbel@mit.edu</td></tr> <tr><td>Devin Caughey</td><td>E53-463</td><td>324-4085</td><td>caughey@mit.edu</td></tr> <tr><td>Nazli Choura</td><td>E53-493</td><td>253-6198</td><td>nchoura@mit.edu</td></tr> <tr><td>Fotini Christia</td><td>E53-417</td><td>324-5595</td><td>cchristia@mit.edu</td></tr> <tr><td>Taylor Fravel</td><td>E40-471</td><td>324-0222</td><td>fravel@mit.edu</td></tr> <tr><td>F. Daniel Hidalgo</td><td>E53-402</td><td>253-8078</td><td>dhidalgo@mit.edu</td></tr> <tr><td>Willard Johnson - Emeritus</td><td>E53-367</td><td>253-2952</td><td>wjohnson@mit.edu</td></tr> <tr><td>In Song Kim</td><td>E53-407</td><td>253-3138</td><td>insong@mit.edu</td></tr> <tr><td>Chappell Lawson</td><td>E53-439</td><td>253-3524</td><td>clawson@mit.edu</td></tr> </tbody> </table> <p>•••</p>	Name	office	phone	email	Regina Bateson	E53-451	324-7336	bateson@mit.edu	Suzanne Berger	E53-451	253-6640	szberger@mit.edu	Adam Berinsky	E53-457	253-8190	berinsky@mit.edu	Donald Blackmer - Emeritus	E53-366	253-3145	blackmer@mit.edu	Andrea Campbell - Department Head	E53-473	452-2295	acampbel@mit.edu	Devin Caughey	E53-463	324-4085	caughey@mit.edu	Nazli Choura	E53-493	253-6198	nchoura@mit.edu	Fotini Christia	E53-417	324-5595	cchristia@mit.edu	Taylor Fravel	E40-471	324-0222	fravel@mit.edu	F. Daniel Hidalgo	E53-402	253-8078	dhidalgo@mit.edu	Willard Johnson - Emeritus	E53-367	253-2952	wjohnson@mit.edu	In Song Kim	E53-407	253-3138	insong@mit.edu	Chappell Lawson	E53-439	253-3524	clawson@mit.edu	<pre> &lt;div id="content" class="clearself"&gt;   &lt;div id="aboutMain"&gt;     &lt;h2&gt;&lt;a href="/polisci/people/faculty/index.html"&gt;PEOPLE&lt;/a&gt; &amp;nbsp; &amp;nbsp;     Faculty&lt;/h2&gt;     &lt;table id="listings"&gt;       &lt;thead&gt;         &lt;tr&gt;           &lt;td class="name"&gt;Name&lt;/td&gt;           &lt;td class="office"&gt;Office&lt;/td&gt;           &lt;td class="phone"&gt;Phone&lt;/td&gt;           &lt;td class="email"&gt;Email&lt;/td&gt;         &lt;/tr&gt;       &lt;/thead&gt;       &lt;tbody&gt;         &lt;tr&gt;           &lt;td class="name"&gt;&lt;a href="/polisci/people/faculty/gina-bateson.html"&gt;Span           class="name"&gt;Regina Bateson&lt;/span&gt;&lt;/a&gt;&lt;/td&gt;           &lt;td class="office"&gt;&lt;span class="office"&gt;E53-453&lt;/span&gt;&lt;/td&gt;           &lt;td class="phone"&gt;324-7336&lt;/td&gt;           &lt;td class="email"&gt;&lt;a href="mailto:bateson@mit.edu"&gt;Span           class="email"&gt;bateson@mit.edu&lt;/span&gt;&lt;/a&gt;&lt;/td&gt;         &lt;/tr&gt;         &lt;tr&gt;           &lt;td class="name"&gt;&lt;a href="/polisci/people/faculty/suzanne-berger.html"&gt;Span           class="name"&gt;Suzanne Berger&lt;/span&gt;&lt;/a&gt;&lt;/td&gt;           &lt;td class="office"&gt;&lt;span class="office"&gt;E53-451&lt;/span&gt;&lt;/td&gt;           &lt;td class="phone"&gt;253-6640&lt;/td&gt;           &lt;td class="email"&gt;&lt;a href="mailto:szberger@mit.edu"&gt;Span           class="email"&gt;szberger@mit.edu&lt;/span&gt;&lt;/a&gt;&lt;/td&gt;         &lt;/tr&gt;         ...       &lt;/tbody&gt;     &lt;/table&gt;   &lt;/div&gt; </pre>
Name	office	phone	email																																																						
Regina Bateson	E53-451	324-7336	bateson@mit.edu																																																						
Suzanne Berger	E53-451	253-6640	szberger@mit.edu																																																						
Adam Berinsky	E53-457	253-8190	berinsky@mit.edu																																																						
Donald Blackmer - Emeritus	E53-366	253-3145	blackmer@mit.edu																																																						
Andrea Campbell - Department Head	E53-473	452-2295	acampbel@mit.edu																																																						
Devin Caughey	E53-463	324-4085	caughey@mit.edu																																																						
Nazli Choura	E53-493	253-6198	nchoura@mit.edu																																																						
Fotini Christia	E53-417	324-5595	cchristia@mit.edu																																																						
Taylor Fravel	E40-471	324-0222	fravel@mit.edu																																																						
F. Daniel Hidalgo	E53-402	253-8078	dhidalgo@mit.edu																																																						
Willard Johnson - Emeritus	E53-367	253-2952	wjohnson@mit.edu																																																						
In Song Kim	E53-407	253-3138	insong@mit.edu																																																						
Chappell Lawson	E53-439	253-3524	clawson@mit.edu																																																						

## Web Scraping : Site structures

Form - "a user interface for modifying web displays."

Display	Source
<p>Roll Call Votes</p> <p>Browse roll call votes in the U.S. Senate and House of Representatives using the filters below.</p> <p><b>TRACK VOTES</b></p> <p>Get an email every time Congress votes on a bill or other matter.</p> <p><b>Filter Votes</b></p> <p>session 2017 (115th Congress)</p> <p>Note: Even-year sessions extend a few days into the next year.</p> <p>chamber <input checked="" type="checkbox"/> All <input type="checkbox"/> House (161 items) <input type="checkbox"/> Senate (90 items)</p> <p>category <input checked="" type="checkbox"/> All <input type="checkbox"/> Procedural (83 items) <input type="checkbox"/> Passage (61 items) <input type="checkbox"/> Amendment (55 items) <input type="checkbox"/> Nomination (21 items) <input type="checkbox"/> Passage under Suspension (16 items) <input type="checkbox"/> Cloture (13 items) <input type="checkbox"/> Unknown Category (2 items)</p> <p>sort by <input type="button" value="Date (Latest First)"/></p>	<pre> &lt;?xml version="1.0"?&gt; &lt;xsl-stylesheet type="text/xsl" href="billres.xsl"?&gt; &lt;!DOCTYPE resolution PUBLIC "-//US Congress//DTD//res.dtd//EN" "res.dtd"&gt; &lt;resolution dss-id="HAC6CFC641CD640B393CFA510D0960B8C" key="H" public-private="public" resolution-stag &lt;metadata xmlns:dc="http://purl.org/dc/elements/1.1/"&gt; &lt;dublinCore&gt;   &lt;dc:title&gt;115 HCON 11 IH: Expressing the sense of Congress that Jerusalem is the capital of Is   &lt;dc:publisher&gt;U.S. House of Representatives&lt;/dc:publisher&gt;   &lt;dc:date&gt;2017-01-23&lt;/dc:date&gt;   &lt;dc:format&gt;text/xml&lt;/dc:format&gt;   &lt;dc:language&gt;EN&lt;/dc:language&gt;   &lt;dc:rights&gt;Pursuant to Title 17 Section 105 of the United States Code, this file is not subjec &lt;/dublinCore&gt; &lt;/metadata&gt; &lt;form&gt;   &lt;distribution-code display="yes"&gt;IV&lt;/distribution-code&gt;   &lt;congress display="yes"&gt;115th CONGRESS&lt;/congress&gt;   &lt;session display="yes"&gt;1st Session&lt;/session&gt;   &lt;legis-type display="yes"&gt;H. CON. RES. 11&lt;/legis-num&gt;   &lt;current-chamber&gt;IN THE HOUSE OF REPRESENTATIVES&lt;/current-chamber&gt;   &lt;action display="yes"&gt;     &lt;action-date&gt;20170123&lt;/action-date&gt;     &lt;action-desc&gt;       &lt;sponsor name-id="B001243"&gt;Mrs. Blackburn&lt;/sponsor&gt; (for herself and &lt;cosponsor name-id="S     &lt;/action-desc&gt;     &lt;action&gt;       &lt;legis-type&gt;CONCURRENT RESOLUTION&lt;/legis-type&gt;       &lt;official-title display="yes"&gt;Expressing the sense of Congress that Jerusalem is the capital of Is       with the location of other United States embassies, the United States       embassy in Israel should be located in Jerusalem.&lt;/official-title&gt;     &lt;/action&gt;   &lt;/form&gt;   &lt;preamble&gt;     &lt;whereas&gt;       &lt;text&gt;Whereas the city of Jerusalem is the seat of Israel&amp;#x2019;s Government, including its P     &lt;/whereas&gt;     &lt;whereas&gt;       &lt;text&gt;Whereas since June 7, 1967, the city of Jerusalem has been an undivided city;&lt;/text&gt;     &lt;/whereas&gt;     &lt;whereas&gt;       &lt;text&gt;Whereas since 1995, it has been Federal law to recognize Jerusalem as the capital of Isr     &lt;/whereas&gt;     &lt;whereas&gt;       &lt;text&gt;Whereas this sense of Congress has no bearing on the final status of Jerusalem as the Un     &lt;/whereas&gt;   &lt;/preamble&gt; </pre>

## Web Scraping : Site structures

XML - "a hierarchical file type to store raw data" (usually displayed in source format)

Display	Source
---------	--------

## H.CON.RES.21

Reaffirming a strong commitment to the United States-Australia alliance relationship.

IN THE HOUSE OF REPRESENTATIVES

FEBRUARY 6, 2017

Mr. ISSA (for himself, Mr. JEA, Mr. ENGLANDER, Ms. KRAUT, Mrs. NATHANSON, Ms. GILDED, Mr. DUCHE, Ms. BERA, Ms. TIA LEE of California, Mr. CASTRO of Texas, Ms. KELLY of Illinois, Mr. SQUIGG, Mr. MURKIN, Mrs. TORRES, Mr. SHRIER, Mr. BRENDAN F. BOYLE of Pennsylvania, Ms. SPIDER, Mr. CONNOLLY, Ms. HANABUSA, Ms. BORDENALO, Mr. HASTINGS, Mr. EVANS, Mr. SMITH of Washington, Mr. ESPALIAR, Mr. COURNEY, Mr. CROWLEY, Mr. Himes, Mr. SCHNEIDER, Ms. TITUS, Mr. COHEN, and Mr. McGOVERN) submitted the following concurrent resolution; which was referred to the Committee on Foreign Affairs

### CONCURRENT RESOLUTION

Reaffirming a strong commitment to the United States-Australia alliance relationship.

Whereas Australia is a vital partner of the United States;

Whereas the United States and Australia share core values as well as deep cultural, security, and people-to-people ties;

Whereas Australia has been a treaty ally of the United States since the signing of the Australia-New Zealand-United States (ANZUS) Treaty in 1951;

Whereas an alliance bond is a sacred vow of friendship and trust, and Australia has always been a faithful and reliable partner to the United States;

Whereas United States-Australia defense and intelligence ties and cooperation are exceptionally close, and Australian forces have fought together with the United States military in every significant conflict since World War I and over 100,000 Australian servicemembers have paid the highest price in the course of their service alongside American allies;

•••

```
<?xml version="1.0"?>
<xm>stylesheet type="text/xml" href="billres.xsl"?>
<xm>declaration public="" uri="http://www.w3.org/2000/svg#><res.dtd">
<resolution dms-id="HCRF9F0D117914B848F4265CCCB70163D" key="H" public-private="public" resolution-stage="Introduced">
<metadata xmlns:rde="http://purl.org/dc/elements/1.1/">
<dublinCore>
<title>H. CON. RES. 21 IH: Reaffirming a strong commitment to the United States-Australia alliance relationship; a concurrent resolution to express the sense of the House of Representatives; introduced in the House of Representatives on February 6, 2017; to support the ANZUS Treaty; to reaffirm the strong alliance between the United States and Australia; to encourage the United States and Australia to continue to work together to promote peace, stability, and prosperity in the Indo-Pacific region; and for other purposes.</title>
<dc:publisher>U.S. House of Representatives</dc:publisher>
<dc:creator>2017-02-06</dc:creator>
<dc:format>text/xml</dc:format>
<dc:language>EN</dc:language>
<dc:rights>Pursuant to Title 17 Section 105 of the United States Code, this file is not subject to copyright laws</dc:rights>
</dublinCore>
</metadata>
<form>
<distribution-code display="yes">HCRF9F0D117914B848F4265CCCB70163D</distribution-code>
<status display="yes">1st Congress</status>
<session display="yes">1st Session</session>
<legis-num display="yes">H. CON. RES. 21</legis-num>
<current-chamber>IN THE HOUSE OF REPRESENTATIVES</current-chamber>
<action display="yes">
<date display="yes">February 6, 2017</date>
<action-desc>
<sponsor name-id="E000179">Mr. Engel</sponsor> (for himself, <cosponsor name-id="I000056">Mr. Issa</cosponsor>)
</action-desc>
</action>
<legis-type>CONCURRENT RESOLUTION</legis-type>
<official-title display="yes">Reaffirming a strong commitment to the United States-Australia alliance relationship; a concurrent resolution to express the sense of the House of Representatives; introduced in the House of Representatives on February 6, 2017; to support the ANZUS Treaty; to reaffirm the strong alliance between the United States and Australia; to encourage the United States and Australia to continue to work together to promote peace, stability, and prosperity in the Indo-Pacific region; and for other purposes.</official-title>
</form>
< preamble>
<text>Whereas Australia is a vital partner of the United States;</text>
</preamble>
<whereas>
<text>Whereas the United States and Australia share core values as well as deep cultural, security, and people-to-people ties;</text>
<whereas>
<text>Whereas Australia has been a treaty ally of the United States since the signing of the Australia-New Zealand-United States (ANZUS) Treaty in 1951;</text>
<whereas>
<text>Whereas an alliance bond is a sacred vow of friendship and trust, and Australia has always been a faithful and reliable partner to the United States;</text>
<whereas>
<text>Whereas United States-Australia defense and intelligence ties and cooperation are exceptionally close, and Australian forces have fought together with the United States military in every significant conflict since World War I and over 100,000 Australian servicemembers have paid the highest price in the course of their service alongside American allies;</text>
<whereas>
<text>Whereas Australia was one of the first countries to commit troops to United States military operations in the Persian Gulf; </text>
<whereas>
<text>Whereas Australia is a close partner of the United States, sharing information essential to the defense of both countries; </text>
<whereas>
<text>Whereas the United States-Australia alliance is an anchor for peace and stability in the Indo-Pacific region; </text>
<whereas>
<text>Whereas Australia has welcomed proposals to reposition United States Marines to maintain Marine forces in the Indo-Pacific region; </text>
</preamble>
<•••>
```

## Web Scraping : Site structures

JSON - "an unstructured file type to store raw data" (the most common web API type)

### Display

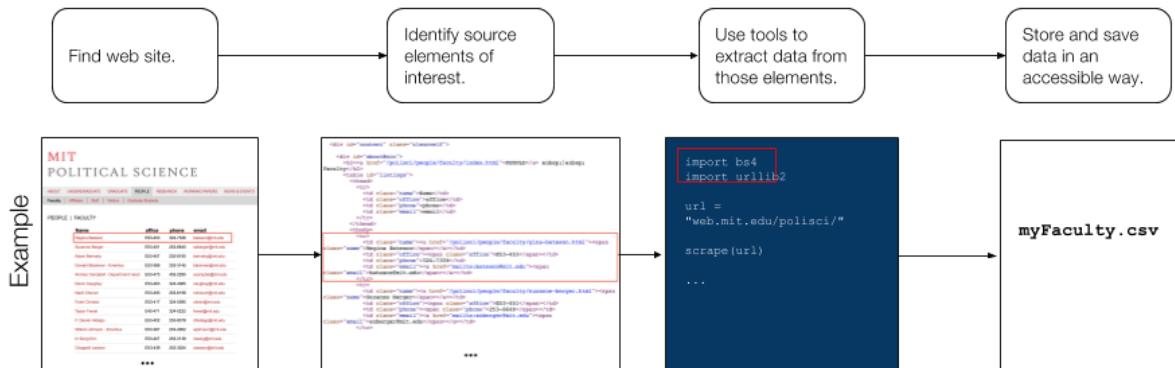
```
{
  "meta": {
    "limit": 441,
    "offset": 0,
    "total_count": 432
  },
  "objects": [
    {
      "created": "2011-03-16T17:49:00",
      "id": 28927519,
      "option": {
        "id": 426366,
        "key": "+",
        "value": "Aye",
        "vote": 1
      },
      "person": {
        "bioguideid": "A000022",
        "birthday": "1942-11-19",
        "cspanid": 1002061,
        "firstname": "Gary",
        "gender": "male",
        "gender_label": "Male",
        "id": 400003,
        "lastname": "Ackerman",
        "link": "https://www.govtrack.us/congress/members/gary_ackerman/400003",
        "middlename": "L.",
        "name": "Rep. Gary Ackerman [D-NY5, 1993-2012]",
        "namemod": ""
      }
    }
  ]
}
```

•••

## Web Scraping : Site structures

## **PDF - "a document presentation file type"**

# Web Scraping : Overview



# Web Scraping : A Case Study in Political Data

- I want to **scrape** online voter records.
  - I want to **scrape** my favorite Senator's wikipedia page.
  - I want to **scrape** Twitter.

... How do I do this in Python?

# Web Scraping : A Case Study in Political Data

First, we need some set up --  
[dev.twitter.com/resources/signup](https://dev.twitter.com/resources/signup)

## Web Scraping : A Case Study in Political Data

(demo)

## Web Scraping : Resource

**Automate the Boring Stuff with Python:** <https://automatetheboringstuff.com/chapter11/>  
[\(https://automatetheboringstuff.com/chapter11/\)](https://automatetheboringstuff.com/chapter11/)

## Web Scraping : Advanced Topics

- Stateful browsing using `mechanize` (useful for forms).
- Text parsing using Regular expressions (useful for finding "patterned" text).
- Natural language processing / text mining (useful for reducing text).
- Streaming / listening applications (useful to get real-time data).
- PDF mining.

## Web Scraping : Applications

Some ideas ...

- Collecting **voter records** for turnout data visualizations.
- Scraping and analyzing city **government careers** pages.
- Aggregating individual **state voting laws** into a database.
- Assessing strength of **public records laws** across different states.

## Intro to Python and Web Scraping

Thanks!

**Contact:** sbarari@mit.edu

*Visit <http://www.mitgovlab.org> (<http://www.mitgovlab.org>)  
to see how MIT GOV/LAB is using web scraping to measure online government transparency.*