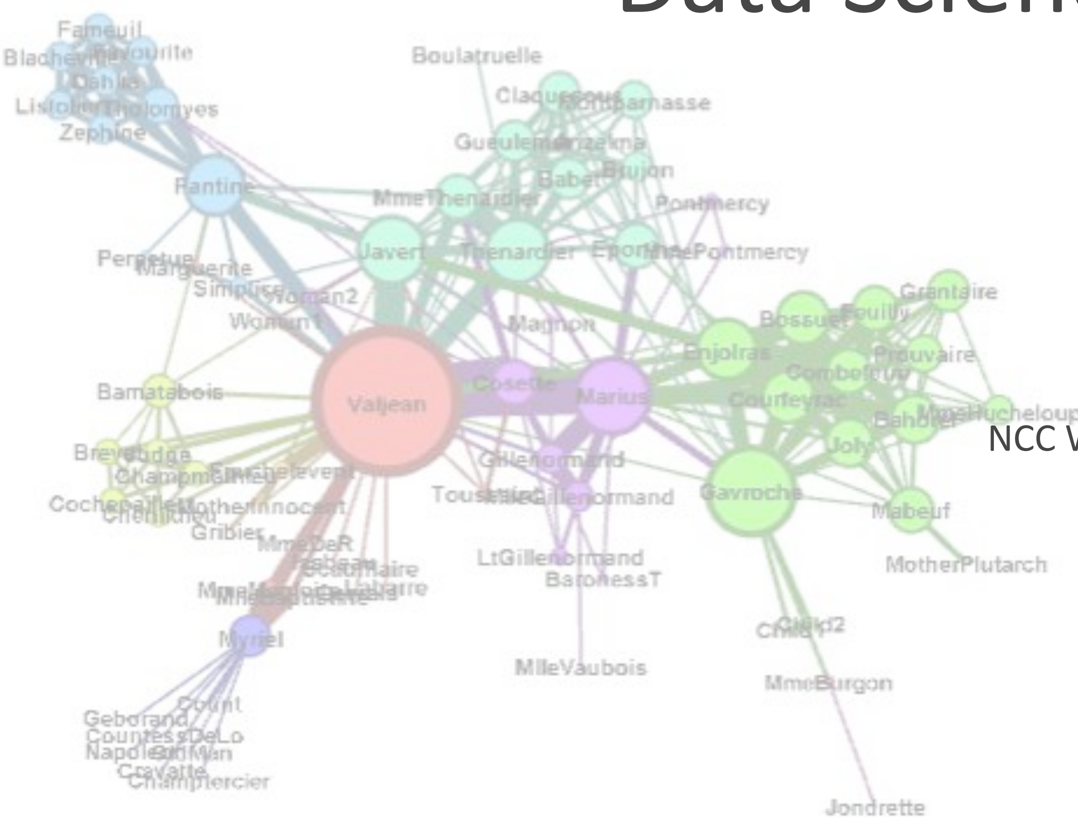



# Data Science & Librarian
# Linked Data

NCC Workshop: Technologies for Librarians
in the Japanese Studies

March 25, 2014

Mariko Honshuku
Librarian for Japanese Law
Harvard Law School Library

# Agenda

## Introduction

Background of today's talk

## Data Science & Librarian

Technologies introduced at DST4L

*Try it! : Web API, Google Table, etc.*

## Linked Data

Intro. Linked Open Data

Linked Open Data in Japan

*Try it! : Hacking Linked Data, etc.*

# Data Science and Librarian

DST4L

# Data Scientist Training for Librarians



> *The ability to take data — to be able to understand it, to process it, to extract value from it, to visualize it, to communicate it — that's going to be a hugely important skill in the next decades.* Hal Varian, McKinsey Quarterly

The Data Scientist Training for Librarians Course aims to upgrade the skill sets of librarians so that they can better serve the data needs of their communities.

The class will be provided with real world use cases and hands-on training covering data extraction, cleansing, sharing and presentation. Members of the course will concentrate on bibliographic data sources and learn programming to the extent that is possible. The group will also look into telling a story with data through various visualization approaches and be introduced to a variety of tools currently used by data scientists.

At the end of the course, participants will develop solutions based on data problems presented at the start of the course and post their work to the DST4L site.




## Recent Posts

- DST4L Feedback Session
- Topic Modeling and Gephi
- Gabriel Florit, Boston Globe Visualizations



4 weeks ago

# Topics included

Clean and analyze data

Data abstraction

Computational thinking

Problem solving skills

Understand data research needs

Agile & continuous, collaborative learnings

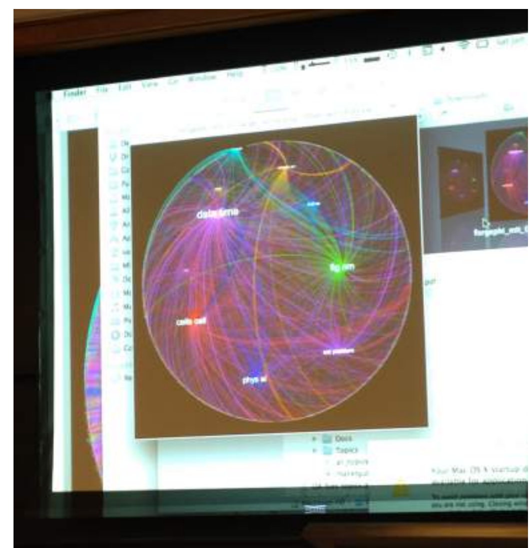
# DST4L

http://swcarpentry.github.io/2013-08-23-harvard/

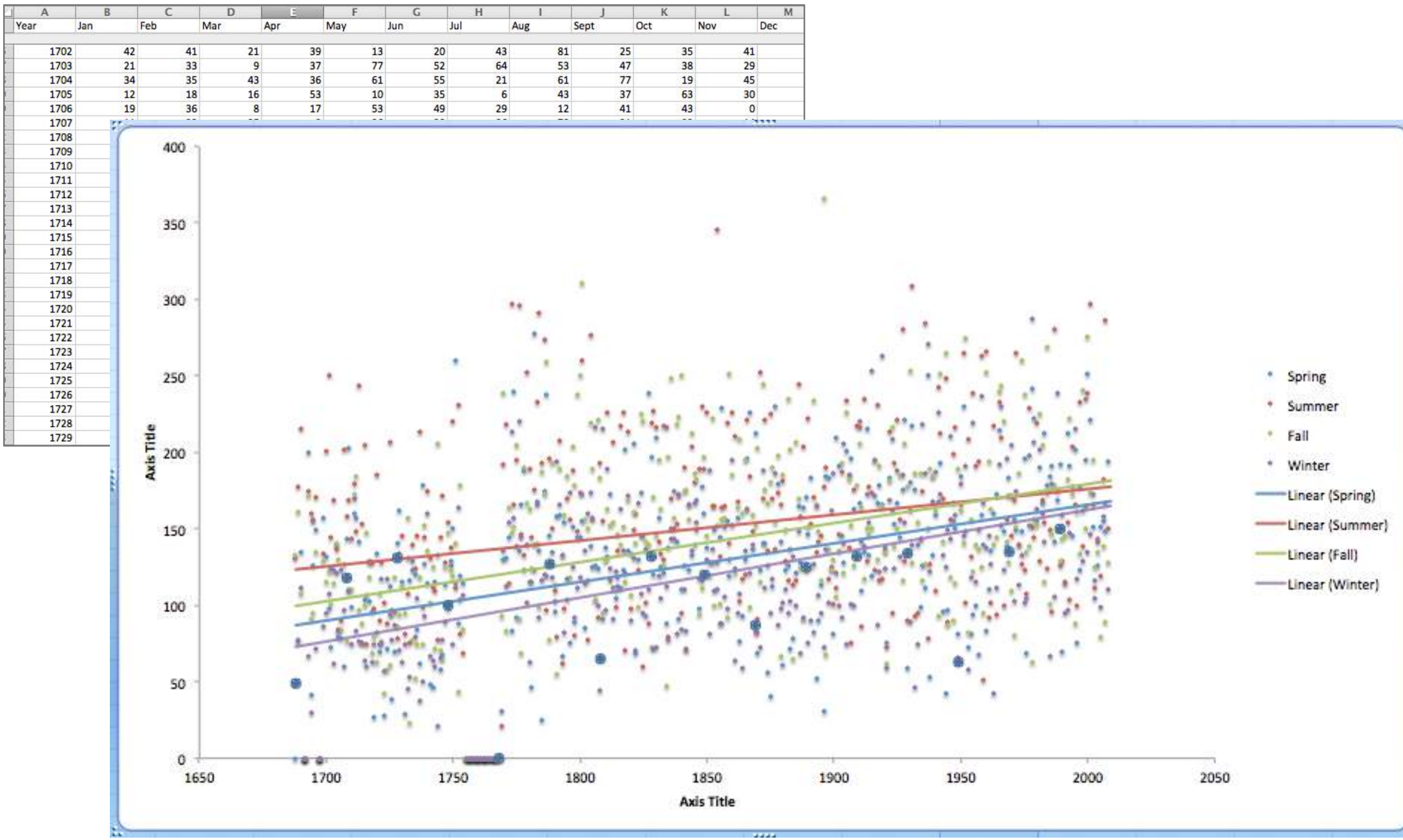
Two days bootcamp in Aug, 2013

Once a week of 3 hours class for 14 weeks

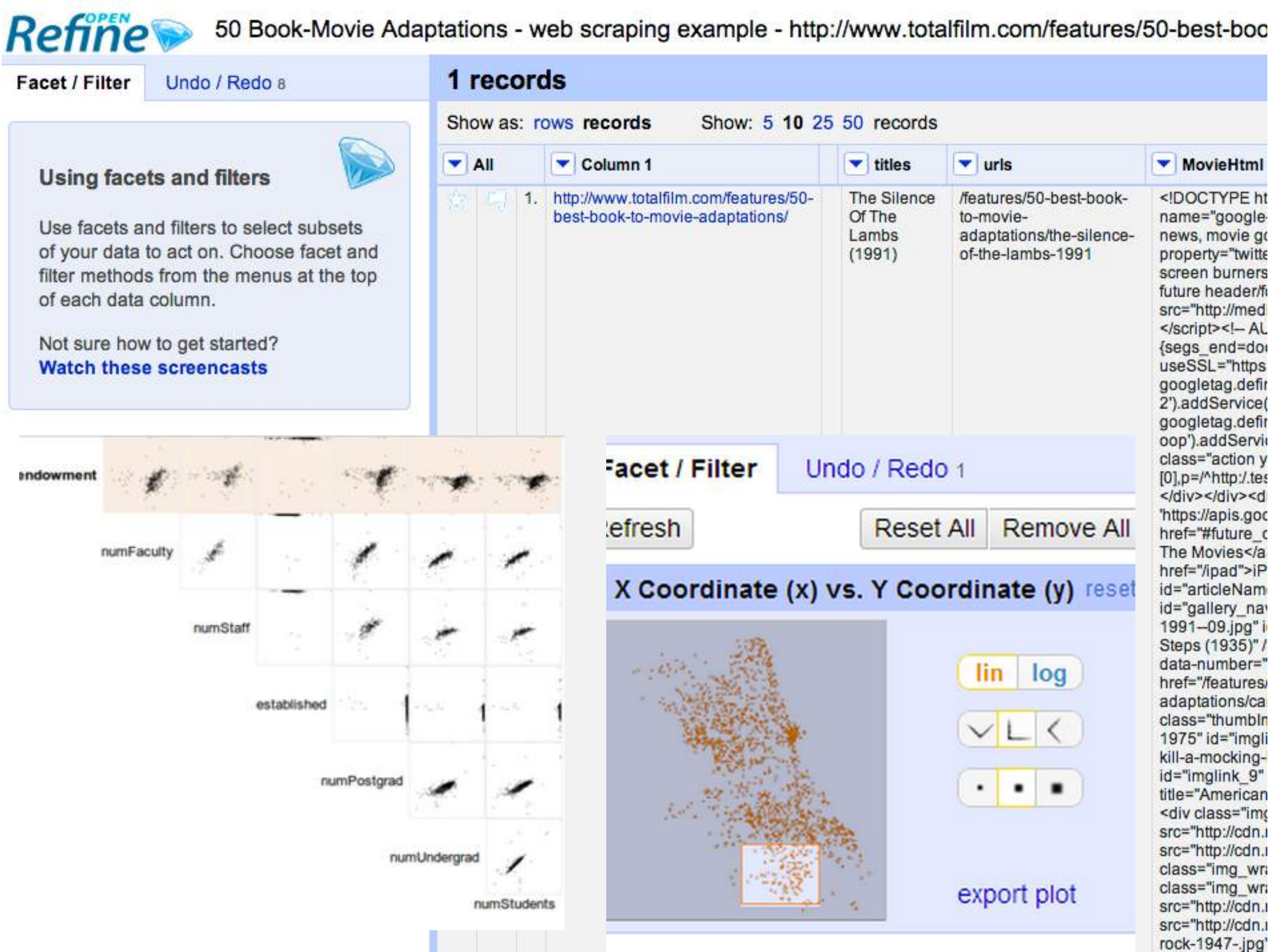
Hackathon at the end



# Excel: Data Cleaning, Manipulations & Graphing

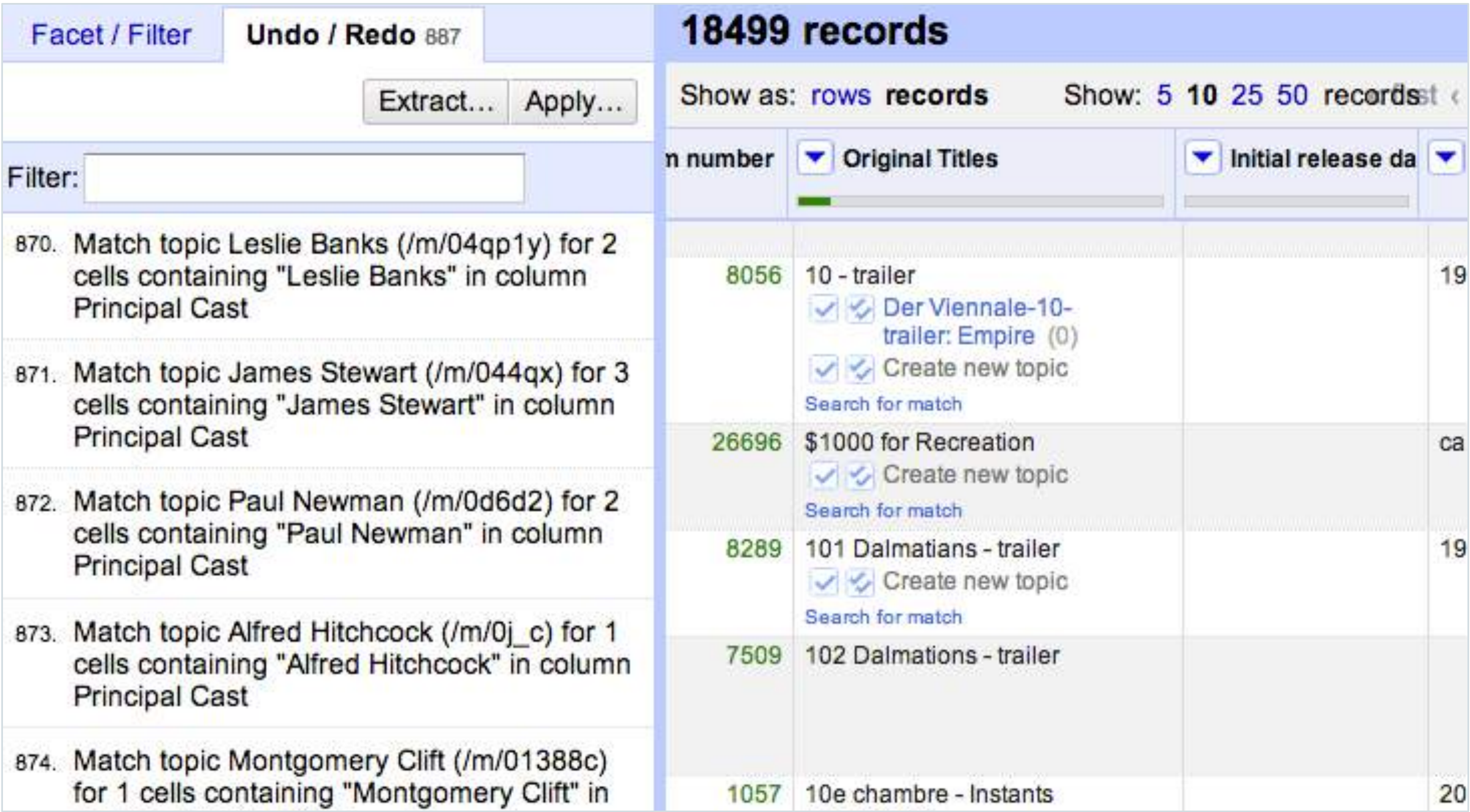


| Year | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sept | Oct | Nov | Dec |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1702 | 42 | 41 | 21 | 39 | 13 | 20 | 43 | 81 | 25 | 35 | 41 | |
| 1703 | 21 | 33 | 9 | 37 | 77 | 52 | 64 | 53 | 47 | 38 | 29 | |
| 1704 | 34 | 35 | 43 | 36 | 61 | 55 | 21 | 61 | 77 | 19 | 45 | |
| 1705 | 12 | 18 | 16 | 53 | 10 | 35 | 6 | 43 | 37 | 63 | 30 | |
| 1706 | 19 | 36 | 8 | 17 | 53 | 49 | 29 | 12 | 41 | 43 | 0 | |
| 1707 | | | | | | | | | | | | |
| 1708 | | | | | | | | | | | | |
| 1709 | | | | | | | | | | | | |
| 1710 | | | | | | | | | | | | |
| 1711 | | | | | | | | | | | | |
| 1712 | | | | | | | | | | | | |
| 1713 | | | | | | | | | | | | |
| 1714 | | | | | | | | | | | | |
| 1715 | | | | | | | | | | | | |
| 1716 | | | | | | | | | | | | |
| 1717 | | | | | | | | | | | | |
| 1718 | | | | | | | | | | | | |
| 1719 | | | | | | | | | | | | |
| 1720 | | | | | | | | | | | | |
| 1721 | | | | | | | | | | | | |
| 1722 | | | | | | | | | | | | |
| 1723 | | | | | | | | | | | | |
| 1724 | | | | | | | | | | | | |
| 1725 | | | | | | | | | | | | |
| 1726 | | | | | | | | | | | | |
| 1727 | | | | | | | | | | | | |
| 1728 | | | | | | | | | | | | |
| 1729 | | | | | | | | | | | | |



OPEN Refine
50 Book-Movie Adaptations - web scraping example - http://www.totalfilm.com/features/50-best-boo
Facet / Filter
Undo / Redo 8
1 records
Show as: rows records
Show: 5 10 25 50 records
All
Column 1
titles
urls
MovieHtml
1.
http://www.totalfilm.com/features/50-best-book-to-movie-adaptations/
The Silence Of The Lambs (1991)
/features/50-best-book-to-movie-adaptations/the-silence-of-the-lambs-1991
Using facets and filters
Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.
Not sure how to get started?
Watch these screencasts
endowment
numFaculty
numStaff
established
numPostgrad
numUndergrad
numStudents
Facet / Filter
Undo / Redo 1
Refresh
Reset All
Remove All
X Coordinate (x) vs. Y Coordinate (y)
reset
lin
log
export plot

# OpenRefine
# reconcile data from freebase.com



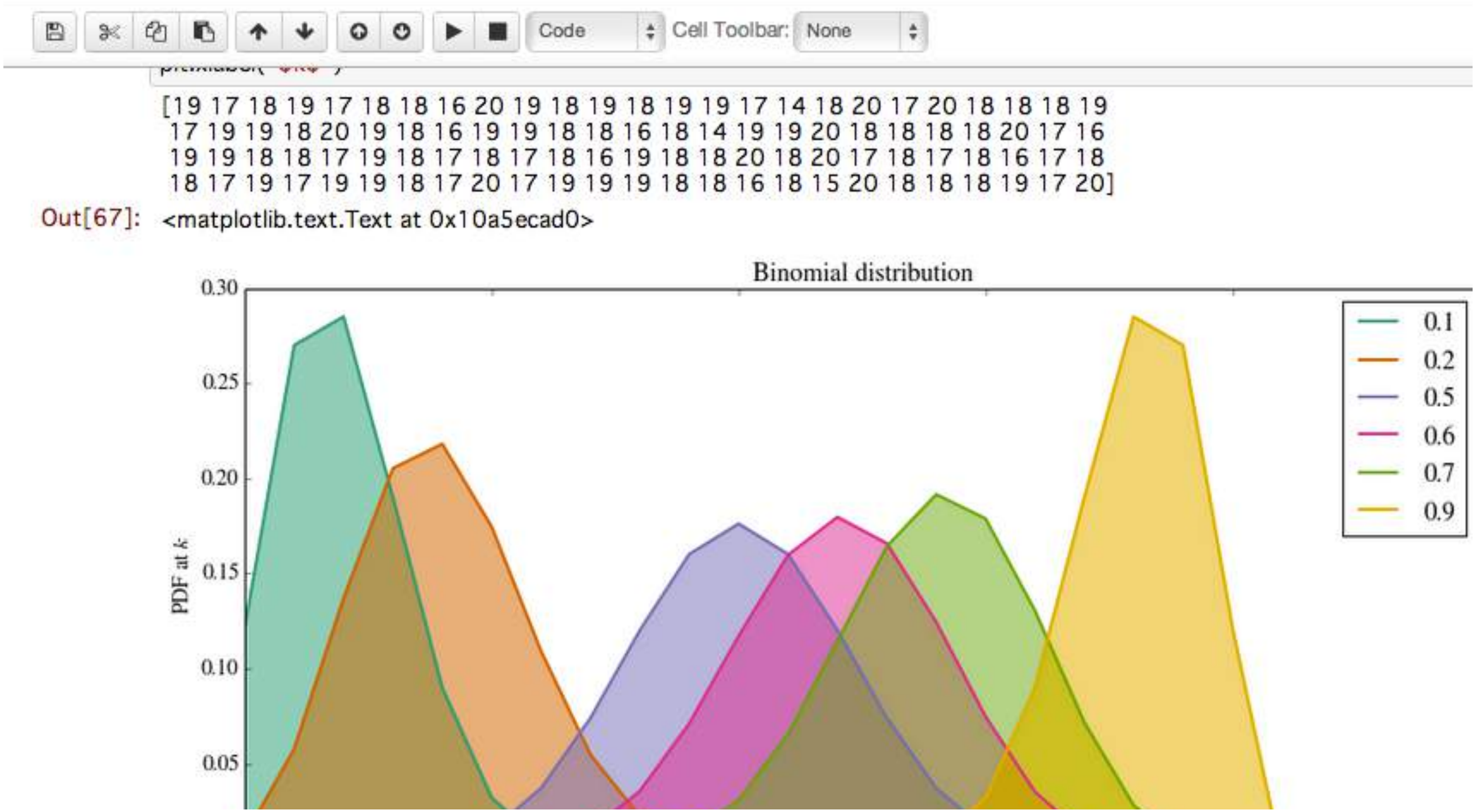
# Python, iPython – Web APIs

```
if authors[author]['works'] > 10:
    url = BASE_URL + author + '.json' # construct the URL that we want
    response = requests.get(url)
    print url
    author_data = response.json()

    for k in ('birth_date','death_date'):
        if k in author_data:
            authors[author][k]=author_data[k]

    # Photos can be accessed in two different forms
    #authors[author]['photo']='http://covers.openlibrary.org/a/
    if 'photos' in author_data and author_data['photos']:
        authors[author]['photo']='http://covers.openlibrary.org/a

    print authors[author]
```

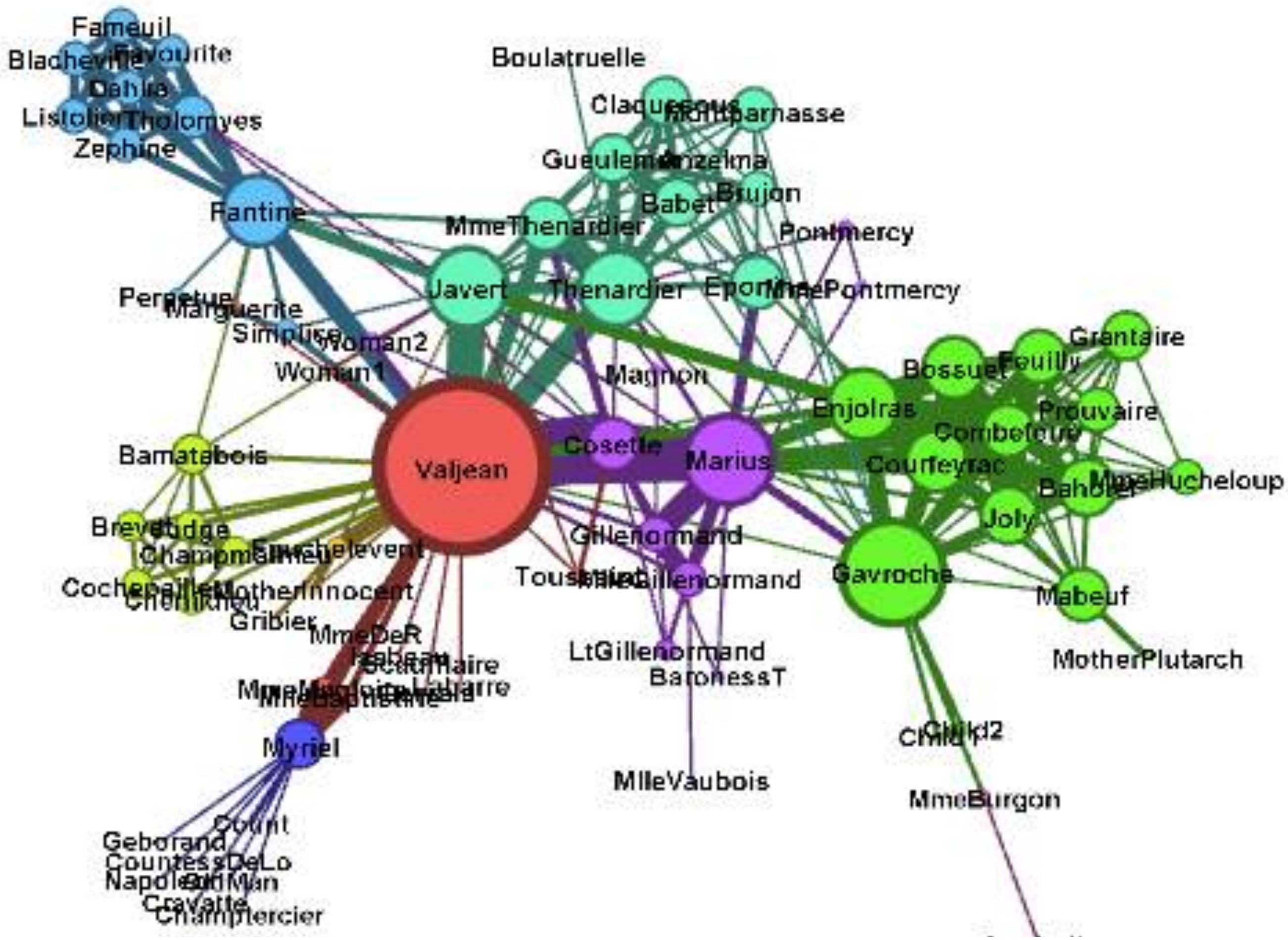
```
http://openlibrary.org/authors/OL786462A.json
{'works': 11, 'name': u'Joannes de Sacro Bosco'}
http://openlibrary.org/authors/OL177839A.json
{'birth_date': u'1791', 'works': 11, 'name': u'Denison Olmsted', 'd
http://openlibrary.org/authors/OL124003A.json
{'birth_date': u'1592', 'works': 11, 'name': u'Pierre Gassendi', 'de
http://openlibrary.org/authors/OL1962966A.json
{'birth_date': u'1786', 'photo': 'http://covers.openlibrary.org/a/i
th_date': u'1853'}
```

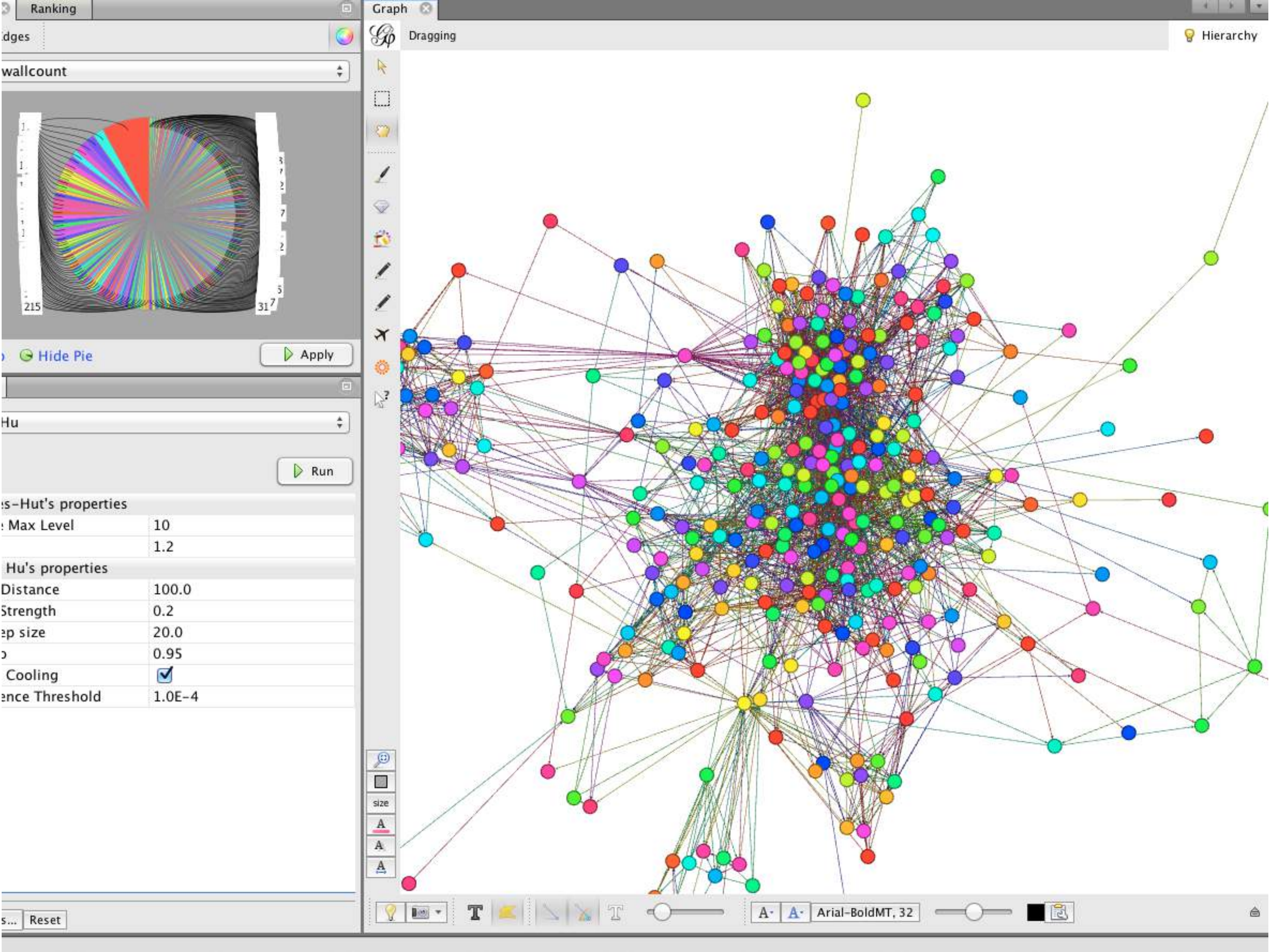

```
{
    name: "Denison Olmsted",
  - created: {
        type: "/type/datetime",
        value: "2008-04-01T03:28:50.625462"
    },
    death_date: "1859",
  - last_modified: {
        type: "/type/datetime",
        value: "2012-06-07T08:33:55.772762"
    },
    latest_revision: 3,
    key: "/authors/OL177839A",
    birth_date: "1791",
    personal_name: "Denison Olmsted",
  - type: {
        key: "/type/author"
    },
    revision: 3
}
```

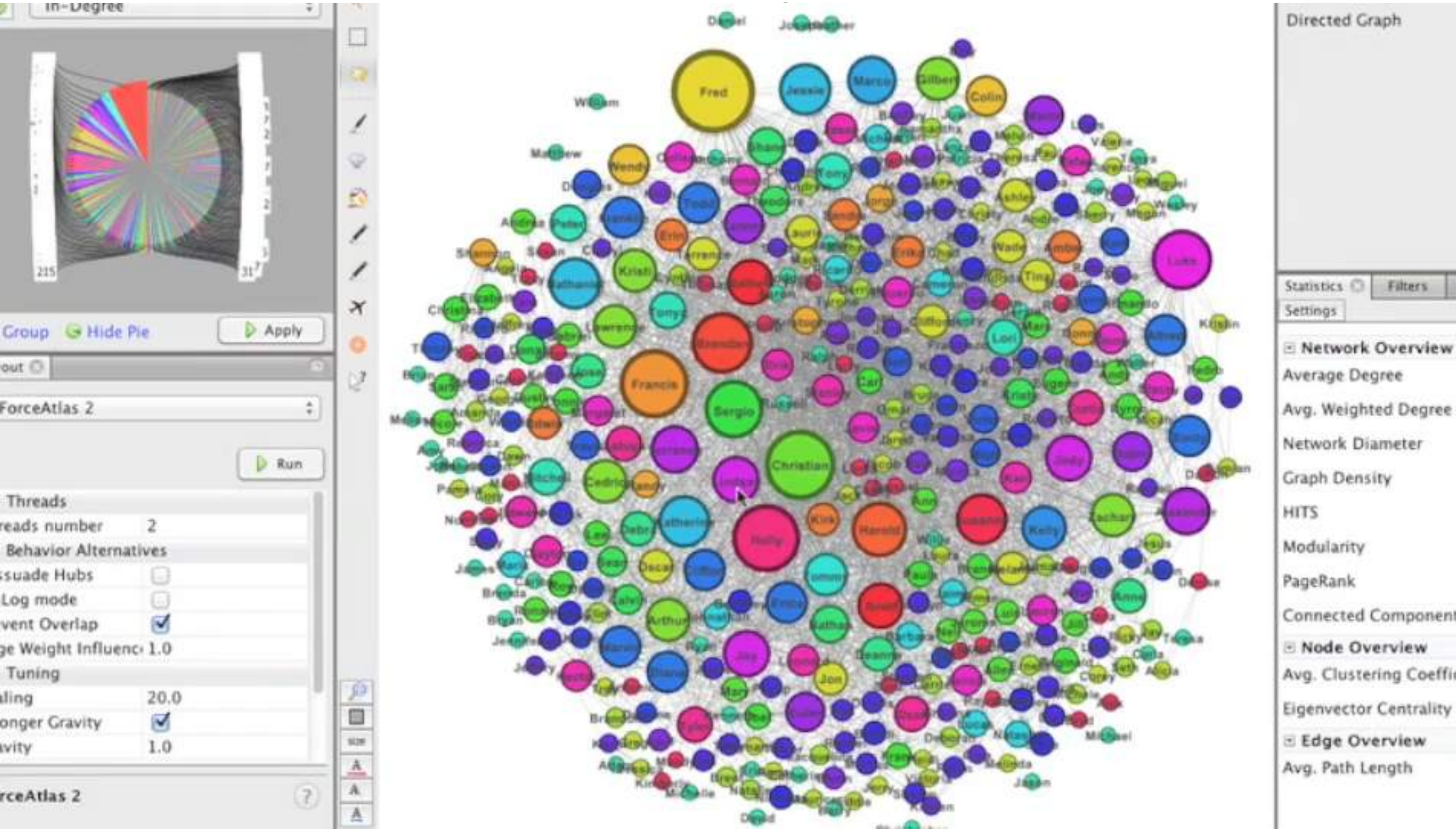
# Python Data Visualization



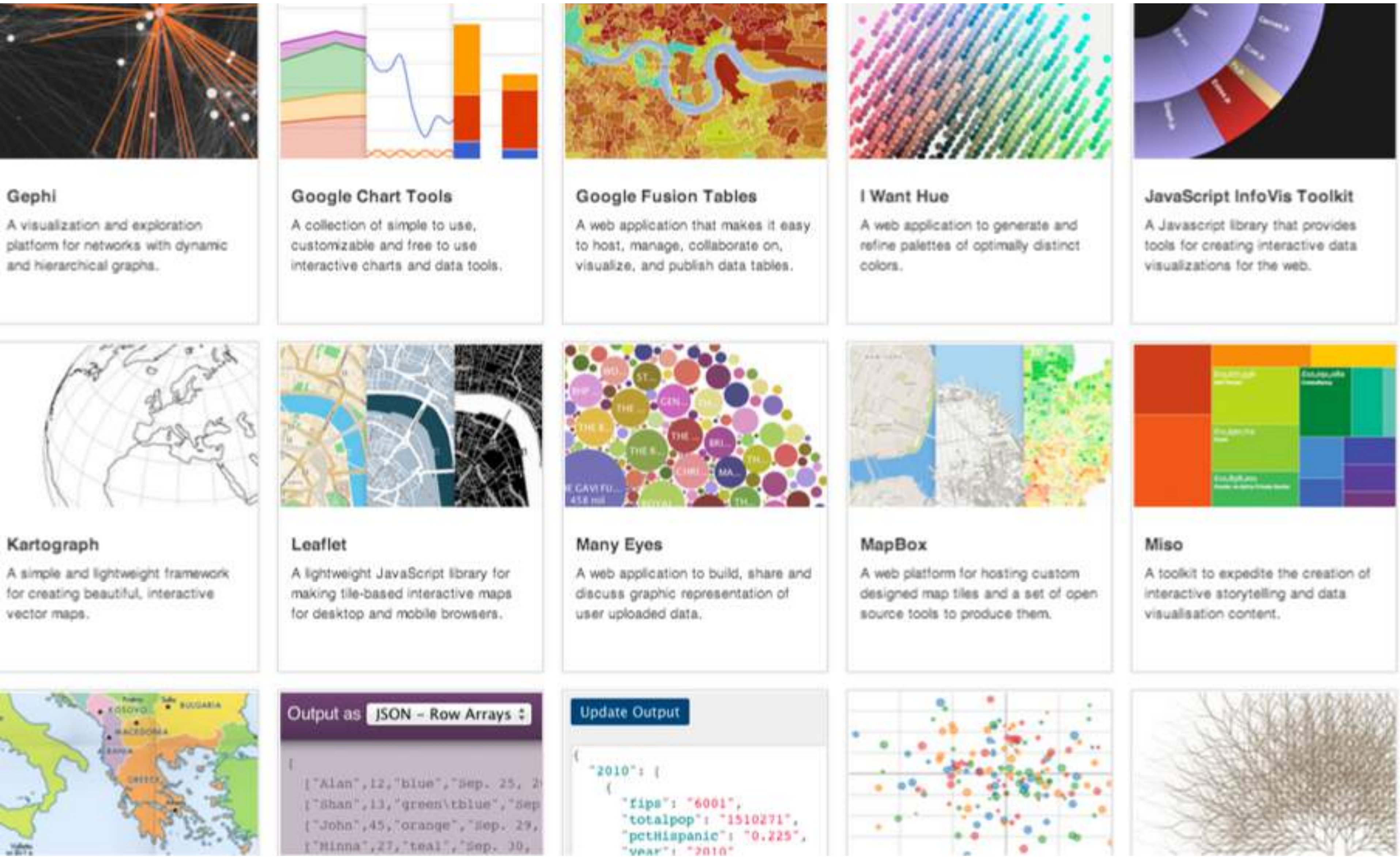
# Gephi

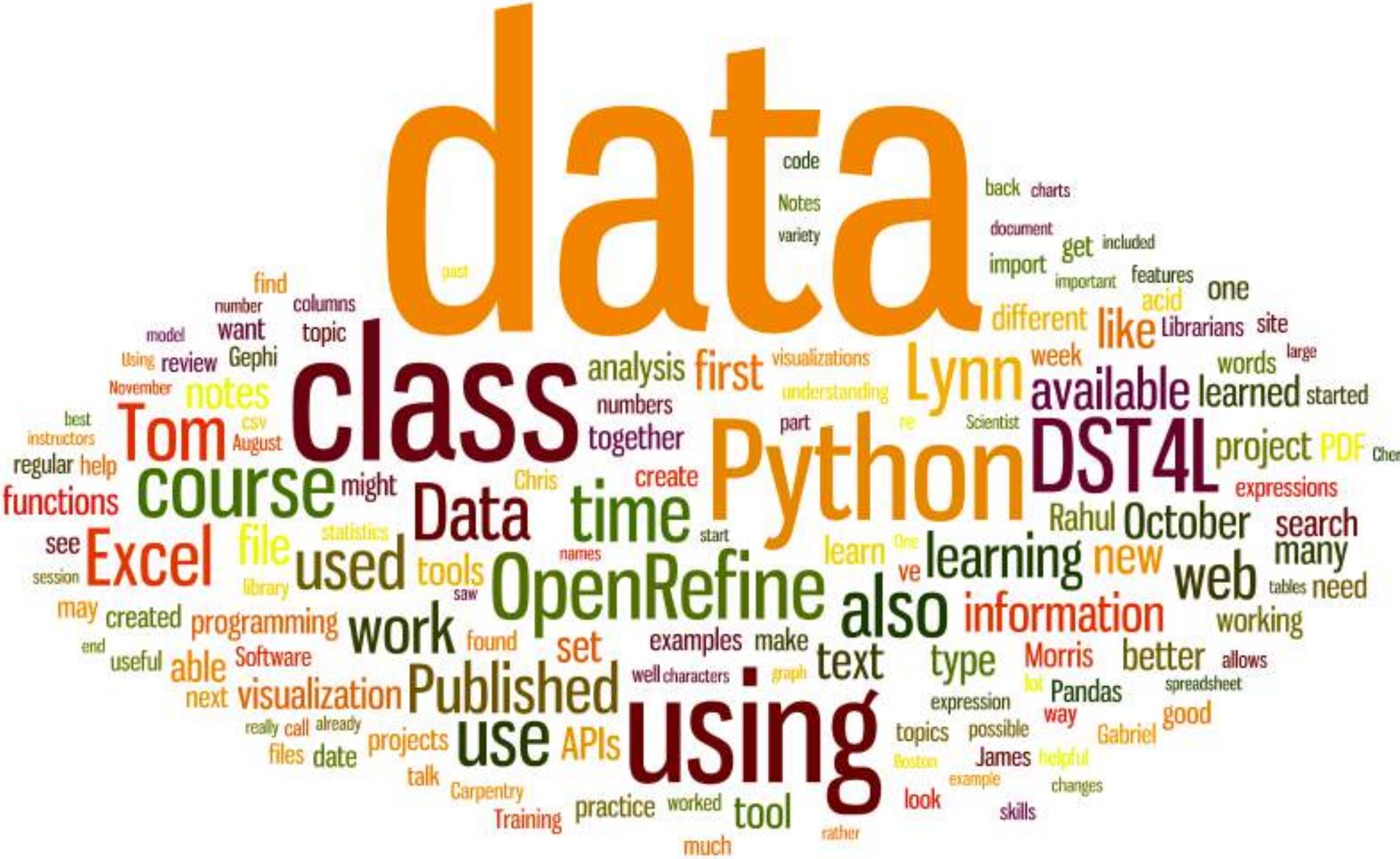

Fameuil
Boulatruelle
Fantine
Javert
Thenardier
MmeThenardier
Babet
Brujon
Pontmercy
Perpetue
Marguerite
Woman2
Woman1
Magnon
Bamatabois
Valjean
Cosette
Marius
Enjolras
Bossuet
Grantaire
Prouvaire
Courfeyrac
Joly
Gavroche
Mabeuf
MotherPlutarch
Gillenormand
LtGillenormand
BaronessT
MlleVaubois
MmeBurgon
Gribier
MmeDeR
Myriel
Geborand
CountessDeLo
Champtercier


Ranking
dges
wallcount
215
31
Hide Pie
Apply
Hu
Run
s-Hut's properties
Max Level
10
1.2
Hu's properties
Distance
100.0
Strength
0.2
ep size
20.0
0.95
Cooling
ence Threshold
1.0E-4
s...
Reset
Graph
Dragging
Hierarchy
size
Arial-BoldMT, 32


In-Degree
Group
Hide Pie
Apply
ForceAtlas 2
Run
Threads
reads number 2
Behavior Alternatives
ssuade Hubs
Log mode
vent Overlap
ge Weight Influenc 1.0
Tuning
aling 20.0
onger Gravity
avity 1.0
rceAtlas 2
Fred
Jessie
Marco
Gilbert
Colin
Wendy
Francis
Sergio
Christian
Harold
Kelly
Luke
Jon
Directed Graph
Statistics
Filters
Settings
Network Overview
Average Degree
Avg. Weighted Degree
Network Diameter
Graph Density
HITS
Modularity
PageRank
Connected Components
Node Overview
Avg. Clustering Coeffici
Eigenvector Centrality
Edge Overview
Avg. Path Length

# datavisualization.ch




data
class
Python
using
DST4L
OpenRefine
course
Tom
Excel
Data
time
Published
used
also
learning
work
use
information
available
learned
Lynn
first
analysis
notes
file
functions
tools
web
text
new
October
search
many
need
project
type
better
Morris
Pandas
set
APIs
programming
visualization
created
able
useful
next
files
date
projects
talk
Carpentry
Training
practice
worked
tool
much
rather
look
skills
changes
example
James
helpful
Boston
topics
possible
expression
way
Gabriel
good
spreadsheet
allows
working
tables
expressions
PDF
started
words
large
site
Librarians
one
like
different
acid
features
important
included
get
import
document
charts
back
code
Notes
variety
past
find
number
columns
model
want
topic
Using
review
Gephi
November
best
instructors
csv
August
regular
help
might
Chris
create
together
numbers
part
re
Scientist
understanding
visualizations
week
Rahul
statistics
start
names
learn
One
ve
see
session
library
saw
may
end
Software
found
examples
make
well
characters
graph
lot
really
call
already

# Let’s Try Some!

1. Download JSON Viewer (Firefox/Chrome)
    1. https://addons.mozilla.org/en-us/firefox/addon/jsonview/
    2. https://chrome.google.com/webstore/detail/jsonview/chklaanhfefbnpoihckbnefhakgolnmc?hl=en
2. Try it on LibraryCloud – Web API
    1. http://librarycloud.harvard.edu/
3. Google Table - Web Scraping
    1. Enter Function (fx)
    2. =importHTML("","table",N)

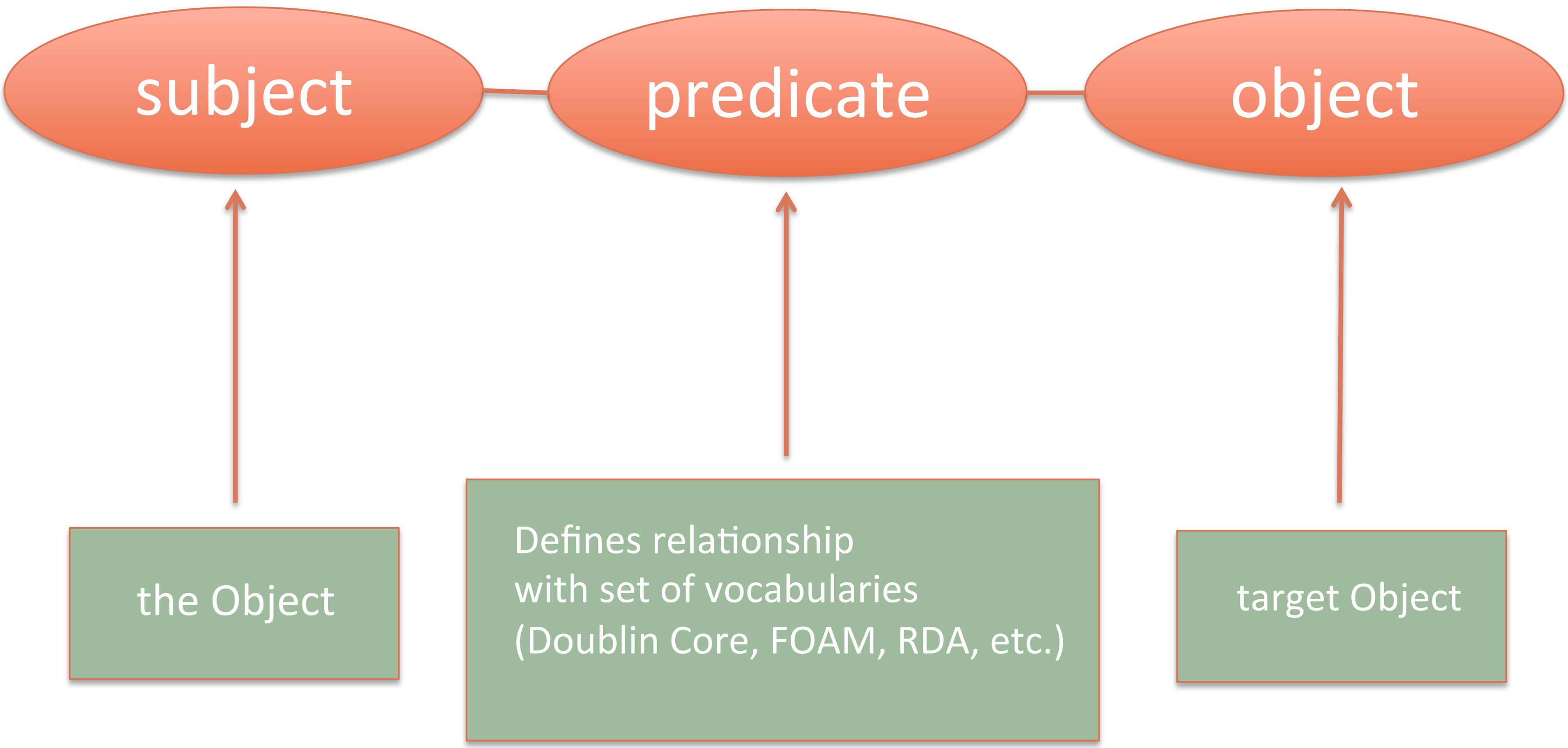
# Linked Open Data

# Linked Data

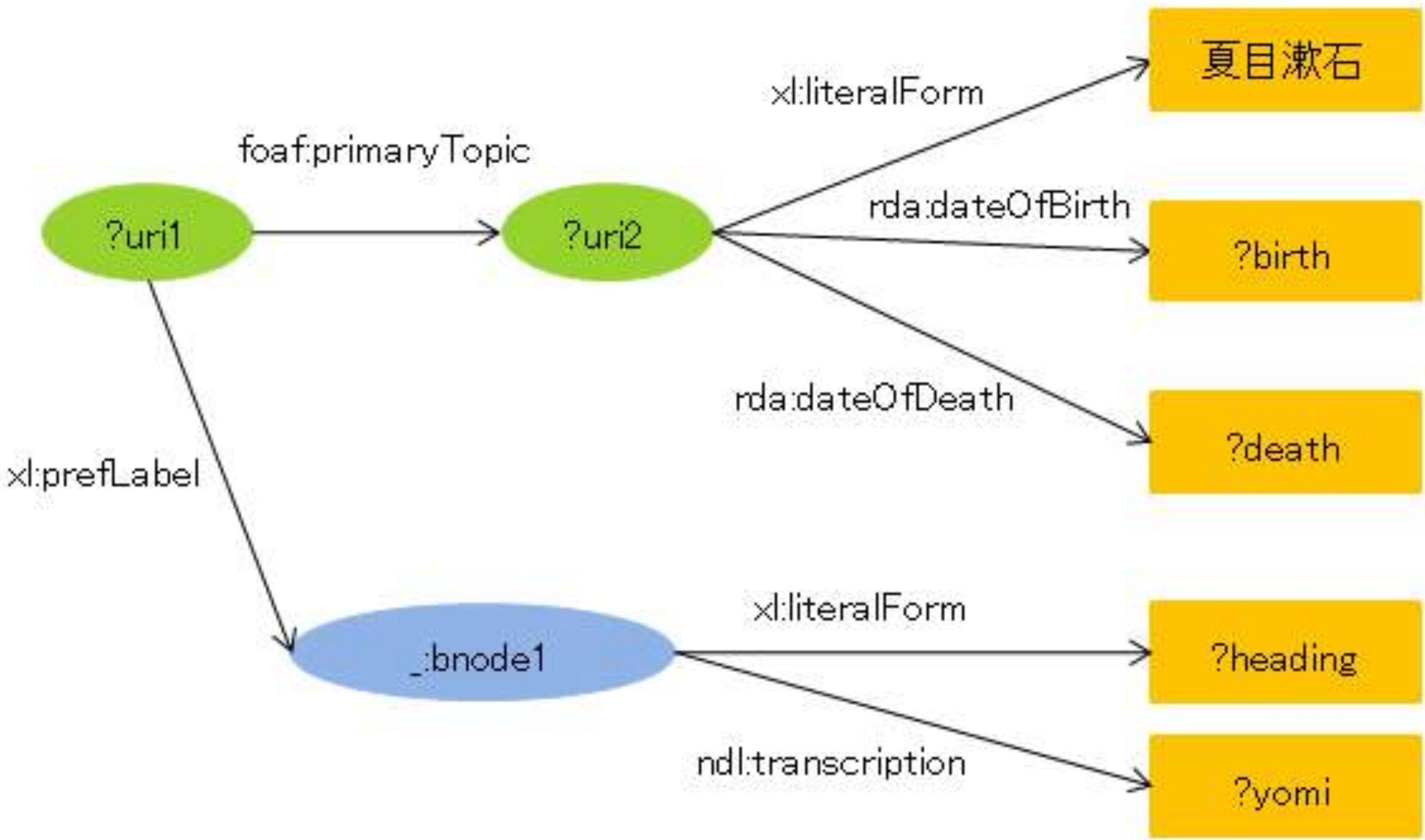
"Linked Data is simply about using the web to create typed links between data from different sources" - *Christian Bizer, Tom Heath, and Tim Berners-Lee, " Linked Data—The story So Far"*

# RDF Triples

**RDF** stands for *Resource Description Framework*
W3C standard for modeling information.



# Web NDL Authorities



# VIAF

Virtual International Authority File

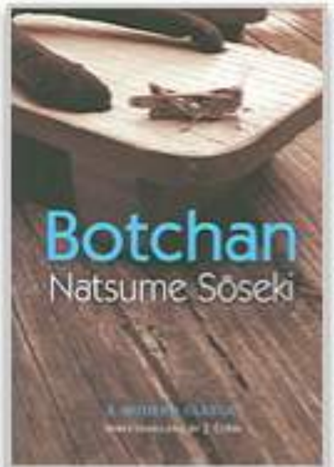
**Search**

Select Field: All Headings

Select Index: All VIAF

Search Terms:

Search




**Natsume, Sōseki, 1867-1916**

**夏目, 漱石, 1867-1916**

**נטסומה, סוסקי, 1867-1916**

**Нацумэ, Сосэки, 1867-1916**

**Soseki, Natsume**

**Natsume, Soseki**

VIAF ID: 56614190 (Personal)

Permalink: http://viaf.org/viaf/56614190

ISNI-test: 0000 0001 2134 0184

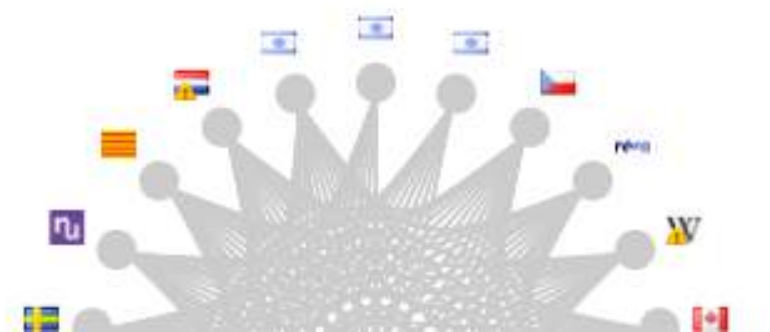
**Preferred Forms**

100 1 _ ‡a Natsume, Sōseki, ‡d 1867-1916

100 1 _ ‡a 夏目, 漱石, ‡d 1867-1916

200 _ | ‡a Natsume ‡b Sōseki ‡f 1867-1916

100 1 _ ‡a Natsume, Sōseki ‡d 1867-1916



datahub
The easy way to get, use and share data

Datasets Organizations About Blog Help Search

/ Organizations / Bibliographic Data / **Web NDL Authorities - ...**

**Web NDL Authorities - National Diet Library of Japan**

Followers
**1**

**Organization**

**Bibliographic Data**

This group comprises open bibliographic datasets according to the Principles on Open Bibliographic Data and

Dataset | Activity Stream | Related

# Web NDL Authorities - National Diet Library of Japan

Name authorities and subject headings of Japan's National Library are provided as Linked Data. Each record is described with such vocabularies as SKOS, SKOS-XL, DC and RDA Elements G2. Records are linked to LCSH, VIAF etc. with SKOS mapping properties where applicable.

## Data and Resources

- API **SPARQL endpoint** — SPARQL endpoint — More information | Go to resource
- HTML **SPARQL query form** — SPARQL query form — Preview | Download
- XML **Example authority record (con neg)** — Example authority record (con neg) — More information | Go to resource
- XML **RDF dump** — RDF dump — More information | Go to resource

authorities · country-japan · deref-vocab · format-dc · format-rdf · format-skos · government · health · library · lld · lod · published-by-producer · subjects

# Web NDL Authorities

国立国会図書館典拠データ検索・提供サービス

キーワード検索　分類記号検索

検索語を入力してください　検索

すべて　名称のみ　普通件名のみ

Web NDL Authoritiesは、国立国会図書館で作成し、 維持管理する典拠データを一元的に検索・提供するサービスです。 サービスの概要については、 「Web NDL Authoritiesについて」 をご覧ください。

最新のお知らせ

- 漢字一文字とひらがな・カタカナが混在している典拠データの検索機能を改修しました (2014年3月6日)
- 【予告】漢字一文字とひらがな・カタカナが混在している典拠データの検索機能を改修します (2014年2月26日)
- 「国立国会図書館典拠データ検索・提供サービス（Web NDL Authorities）」の機能を拡張しました (2012年12月13日)
- 【予告】「国立国会図書館典拠データ検索・提供サービス（Web NDL Authorities）」のメタデータ項目の変更について (2012年12月7日)
- 「国立国会図書館典拠データ検索・提供サービス（Web NDL Authorities）」の本格サービスを開始しました (2012年1月6日)

# Web NDL Authorities

国立国会図書館典拠データ検索・提供サービス

◀ 検索結果に戻る

## インターネット

| 詳細情報 | グラフィカル表示 |
|---|---|

| | |
|---|---|
| **ID** | 00841024 |
| **標目** (xl:prefLabel) | インターネット<br>インターネット |
| **同義語** (xl:altLabel) | Internet |
| **注記** (skos:note) | 当件名新設 (2001年5月25日) 以前の整理では，「データ伝送」と「通信網」を合わ |
| **上位語** (skos:broader) | コンピュータネットワーク |
| **下位語** (skos:narrower) | バーチャルプライベートネットワーク; イントラネット; セマンティックウェブ |
| **関連語** (skos:related) | インターネット放送; チャット; 情報活動 (インターネット); IP電話; 検索エンジン; イ<br>グ; ネットワークゲーム; ホームページ; インターネット広告; インターネットプロバイ<br>ション; ネット犯罪; ポッドキャスティング; ドメイン名; インターネット依存症 |
| **分類記号** (skos:relatedMatch) | DK341 (NDLC); ND633 (NDLC); 547.483 (NDC9); 694.5 (NDC9) |
| **関連リンク** | sh92002816 (LCSH) |

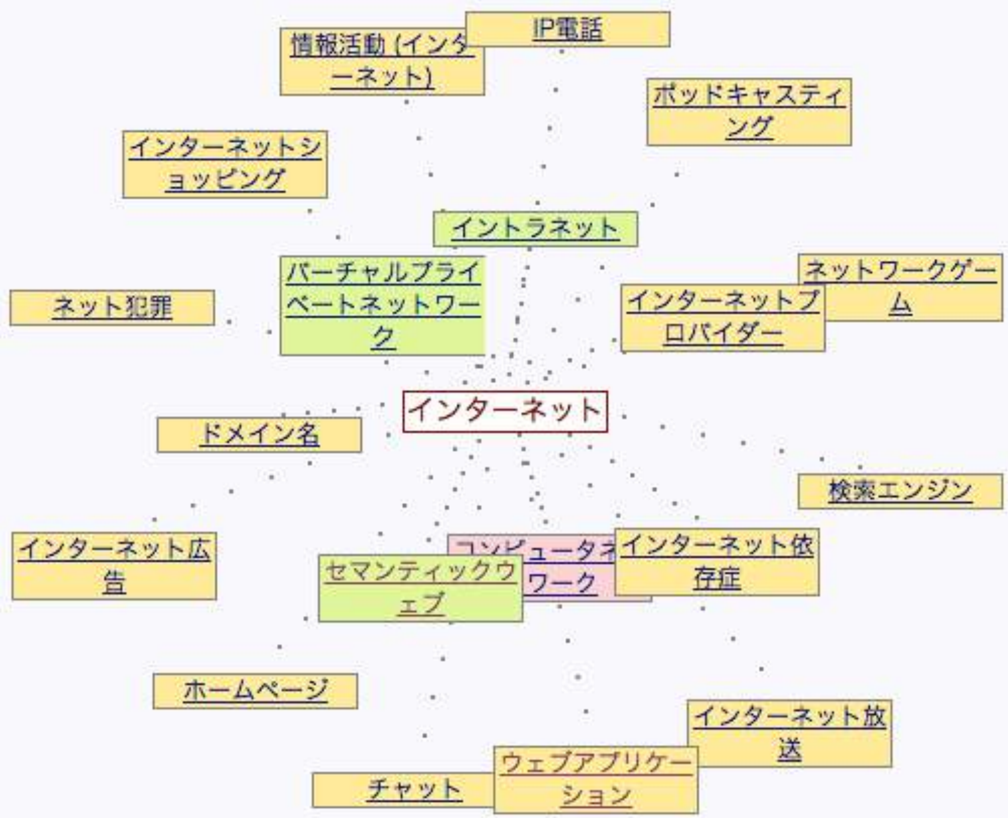
# Web NDL Authorities

国立国会図書館典拠データ検索・提供サービス

◀ 検索結果に戻る

## インターネット

詳細情報 | グラフィカル表示



about this App 誰でもかんたんSPARQL実行 useful un-useful 2 useful

Preview Input Data ReadMe Snapshots

# 誰でもかんたんSPARQL実行

使用するエンドポイントを選択: 国立国会図書館典拠データ検索・提供サービス（Web NDL Authorities）

| | |
|---|---|
| **エンドポイントの説明** | 国立国会図書館が維持管理する典拠データを一元的に検索・提供するサービス。 |
| **エンドポイントのURL** | http://id.ndl.go.jp/auth/ndla |
| **制作者** | 国立国会図書館 |
| **ライセンス** | http://iss.ndl.go.jp/ndla/use/ |
| **関連情報** | http://id.ndl.go.jp/auth/ndla/ |

```
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>  PREFIX rdfs:
<http://www.w3.org/2000/01/rdf-schema#>  PREFIX rda:
<http://RDVocab.info/ElementsGr2/>  PREFIX foaf: <http://xmlns.com/foaf/0.1/>  PREFIX xl:
<http://www.w3.org/2008/05/skos-xl#>  PREFIX ndl: <http://ndl.go.jp/dcndl/terms/>
SELECT * WHERE {  ?uri1 foaf:primaryTopic ?uri2.  ?uri1 xl:prefLabel [ xl:literalForm ?heading;
ndl:transcription ?yomi ].  ?uri2 rda:dateOfBirth ?birth.  ?uri2 rda:dateOfDeath ?death.  ?uri2
foaf:name "夏目漱石".  FILTER (lang(?yomi) = 'ja-Kana').  }
```

クエリ例 1 クエリ例 2 クエリ例 3 実行

「夏目漱石」の典拠情報URI、名称実体URI、生年、没年、標目、標目カナヨミを調べる。

# Let’s Try SPARQL!

# IT Strategic Headquarters

高度情報通信ネットワーク社会推進戦略本部（ＩＴ総合戦略本部）

http://japan.kantei.go.jp/policy/it/indexe.html

Open Government Strategy
電子行政オープンデータ戦略 2012.7

Declaration to be the World's Most Advanced IT Nation
世界最先端IT国家創造宣言 2013.6

# Open DATA METI Beta
Empowering Economy, Trade and Industry



This website, Open DATA METI (Beta), was launched in January 2013, as an experimental site to examine our open data policy.We will make this site more useful by receiving opinions of enterprises, NPOs, and individuals, and also reflect your opinions to progress open data policy of Japanese government with consideration of personal information. Therefore, we organized group "DATA METI partners" to receive your opinions and demands. Summary of this group and application guidance is written in this page (Japanese only).

We will modify this website based on your comments. Please understand that any modifications on this website without precaution is possible, and we cannot take responsibility for any damage caused by these modifications.

## Updated datasets

- 経済産業省生産動態統計調査(2012年) （2014-03-18 10:50）
- 規模別製造工業生産指数(2010年) （2014-03-18 10:50）
- 規模別製造工業生産指数(2011年) （2014-03-18 10:50）
- 規模別製造工業生産指数(2012年) （2014-03-18 10:50）
- 規模別製造工業生産指数(2013年) （2014-03-18 10:50）
- 中小企業実態基本調査(2003年度) 調査 （2014-03-18 10:50）

## What's new

- 平成25年度IT融合フォーラム公共データWG（第２回）の資料の公開について （2014-03-05 18:14）
- 平成25年度IT融合フォーラム公共データWG（第２回）への傍聴者の募集について （2014-02-26 18:53）
- オープンデータユースケースコンテストの結果を公開しました！ （2014-02-14 11:25）
- オープンガバメントアイディアボックス2014を開設しました！ （2014-02-06 21:20）
- Open DATA METIの機能拡充のお知らせPart3 （2014-01-17 17:28）
- データカタログサイト試行版　DATA.GO.JPの公開について （2013-12-20 13:28）

## このサイトについて


このサイトは、「データカタログサイト」の試行版として、各府省の保有データをオープンデータとして利用できる場をつくり、データの提供側・利用側双方にオープンデータのイメージを分かりやすく示すことを目的としています。

これにより、オープンデータの利用を促進しその利用事例の収集に役立てるとともに、試行版に対する利用者等の意見をもとに、本格的な「データカタログサイト」の整備に向けた改善点の検討を行います。

登録しているデータについては、「電子行政オープンデータ推進のためのロードマップ」において重点分野とされた白書、防災・減災情報、地理空間情報、人の移動に関する情報、予算・決算・調達情報を中心に登録しております。

## 新着情報

| | |
|---|---|
| 「オープンデータ・アプリコンテスト」「勝手表彰」審査結果発表 | 2014/03/17 |
| システムの定期メンテナンスのお知らせ | 2014/03/17 |
| ９４４件のデータセットを新規登録しました。 | 2014/03/06 |

welcome!
openlabs.go.jp
開かれた行政を目指す実験サイト

このサイトを検索




## オープンガバメントラボへようこそ

このサイトは、日本におけるインターネットを活用して開かれた政府（オープンガバメント）の実現を目指し、さまざまな実証を行っているサイトです。オープンガバメントに関するさまざまな情報の収集、提供も行っています。

## お知らせ

OpenDATAMETIアイディアソン 本アイディアソンでは、経済産業省をはじめ、政府が持つデータの有効な活用方法について参加者全員で考えます。「官民の情報連携」と「中小企業情報」の２つを大きなテーマとし、経済産業省の政策担当者とともに情報の利活用のアイディア、政策提言を考える場にしていきたいと思います。オープンデータに関する事業をよりポジティブに進めるために、行政職員も積極的に参加し、これまでとは違ったアイディアを

東日本大震災
復旧・復興支援制度
検索サービス

ネットアクション
開かれた復興へ
あたらしい日本へ

復旧・復興支援
サイト/アプリ等調査事業
投票・推薦はこちら

公共機関
ソーシャルメディアポータル

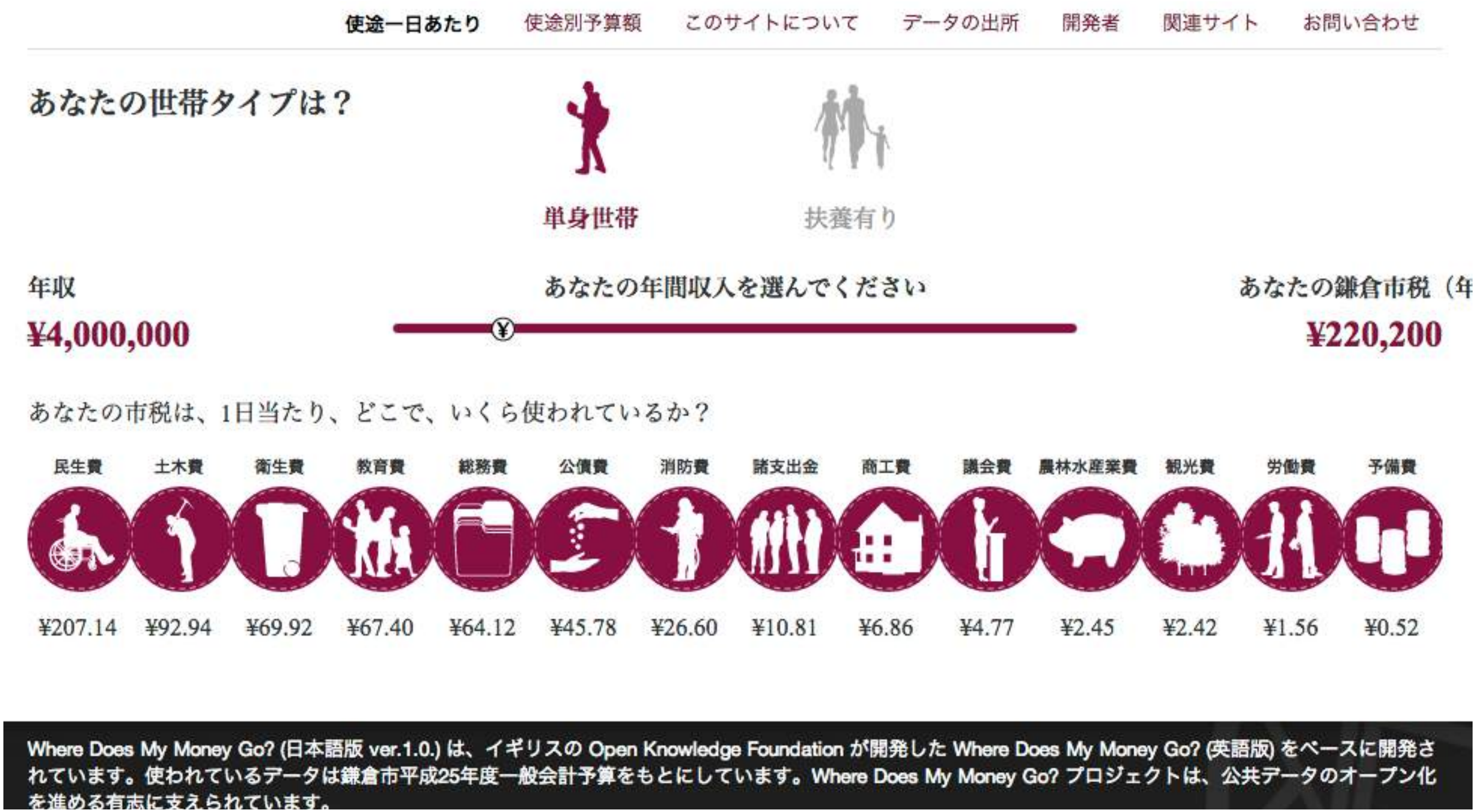
オープンガバメントWiki
国内外のオープンガバメント情報の集積場

IdeaBox Archives
過去行われたアイディアボックス

# Open Linked Data in Japan

- 2011
  - Linkdata.org
- 2012
  - Open Knowledge Foundation Japan
  - Open Data Promotion Consortium
  - linkedopendata.jp
- 2013
  - Open Data METI
  - Data.go.jp

# Where Does My Money Go?



June 30, 2012 — July 1, 2012

# GLOCOMオープンデータ活用ハッカソンで横浜版がローンチ

GLOCOM オープンデータ活用ハッカソンにて、川島氏のアイデアを元に横浜版を立ちあげ。ハッカソンでは最優秀賞を受賞。

Source: レポート：GLOCOMオープンデータ活用ハッカソン

JUNE 30, 2012

横浜市版

GLOCOMオープンデータ活用ハッカソンで横浜版がローンチ

横浜市版

JUNE 30

JULY

JULY 4

JULY 8

Japanese

# OPEN DATA
PROMOTION CONSORTIUM

OUTLINE | GENERAL MEETING AND COMMITTEES | EVENT | NEWS | MEMBERSHIP GUIDE


## About the Consortium

Open Data Promotion Consortium, a group established in 2012 as a base for industry-government-academia joint activities, aims to improve the environment for promoting the distribution of open data.

» Outline of the Consortium » About Committees » Membership Guide

Open Data Symposium

# What We Should Learn from Global Trends toward the Most Advanced Open Data Society

Monday 9 December 2013

| | |
|---|---|
| Venue | The Ito Hall, the University of Tokyo |
| Organizers | Open Data Promotion Consortium, Ministry of Internal Affairs and Communications |
| Supporters | Ministry of Economy, Trade and Industry, Ministry of Land Infrastructure and Transport, KEIDANREN, ASP-SaaS-Cloud Consortium, National Diet Library |

## Program

| Hour | Length | Session | Speaker |
|---|---|---|---|
| 13:00 | | Registration | |
| 13:30 | | Opening | |
| 13:30-13:35 | 5 minutes | Welcome | **Masahito Fujikawa** Vice-Minister for Internal Affairs and Communications |
| 13:35-13:50 | 15 minutes | Welcome and Short-Presentation | **Hiroshi Komiyama**, Chairman of Open Data Promotion Consortium |

NIKKEI Language

# Nikkei Open Data Portal

オープンデータ、データジャーナリズムに関する情報と実験のサイトです。

View details ›

Home | Article | About | Links

input keyword Search

## アプリコンテスト最優秀賞は「花粉くん」　情報の受け手視点でデータ活用

オープンデータ流通推進コンソーシアムと総務省は３月13日、都内で「オープンデータ・アプリコンテスト」の表彰式を行った。

(2014/03/13)

View details »

### TAG

データジャーナリズム　リポート　ニュース　インタビュー　事例

### BACK NUMBER

- アプリコンテスト最優秀賞は「花粉くん」　情報の受け手視点でデータ活用
- オープンデータで「食の日本地図」　最新版にアップデート


LinkData.org
Contact
Login
Languag
LinkData.org
オープンデータ活用支援プラットフォーム
Data
Application
Idea
City
Link Data
See or Publish Data
App LinkData
See or Publish Apps
Idea LinkData
See or Publish Idea
CityData
See by Cities
Search Specialists
Enter keyword...
Search

App LinkData

Publish your application as JavaScript that mashes up the Linked Open Data

HOME | Tutorial | Create App | Fork App | Published Apps

## Sections



## Introduction

LinkDataApp provides the platform for creating and publishing JavaScript Applications.

You can create your own Applications using LinkData works in **LinkData.org** as input data.