Wikimedia Public Resources

Diego Sáez Trumper - Nov. 2018

diego@wikimedia.org

WIKIMEDIA FOUNDATION

Open access policy

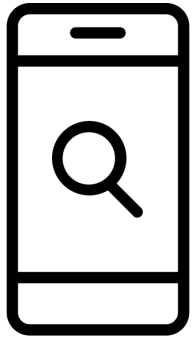
The Wikimedia Foundation's mission is to disseminate open knowledge effectively and globally. In keeping with this mission, the Wikimedia Foundation supports research in areas that benefit the Wikimedia community. We aim to make any work produced with our support openly available to the public and reusable on Wikimedia projects.

Minimalist user data collection

Revision histories are public data

No 3rd-party collection or sharing of user data

Private user data removed after 90 days



What can I do?

- NLP
  - Topic Models
  - Word Embeddings
  - Quality
- Graphs
  - Page links
  - Users graph
- Geolocation
- Socio-technical
  - User Influence
  - Cultural Differences
  - Online Deliberation
  - Articles Popularity
- Multimedia
  - Audio, Video, Image Analysis
- And more...

WIKIMEDIA FOUNDATION

Research Resources

- Statistics
- Dumps
- MediaWiki Utilities
- Wikimedia API
- Page Views
- SQL Replicas / Quarry
- Clicks
- Event Stream
- Wikidata
- Commons
- ORES
- PAWS

WIKIMEDIA
FOUNDATION

New York University

From Wikipedia, the free encyclopedia

Coordinates: 40°43′48″N 73°59′42″W

This article is about the private institution founded in 1831. For other and similar uses, see University of New York.
"Nyu" redirects here. For the district in Japan, see Nyū District, Fukui. For the Elfen Lied character, see List of Elfen Lied characters § Nyu.

New York University (**NYU**) is a private nonprofit research university based in New York City. Founded in 1831, NYU's primary campus is in Greenwich Village with other campuses throughout New York City.[12][13] NYU students can also study abroad at its degree-granting campuses in NYU Abu Dhabi and NYU Shanghai, as well as its 11 academic centers in Accra, Berlin, Buenos Aires, Florence, London, Madrid, Paris, Prague, Sydney, Tel Aviv, and Washington, D.C.[14][15]

In 2018, NYU was ranked amongst the top 30 universities internationally by the *Academic Ranking of World Universities*, *Times Higher Education World University Rankings*, and *U.S. News & World Report*.[16][17][18] For the

‹ The template *Infobox university* is being considered for merging. ›

New York University

Latin: *Universitas Neo Eboracensis*

Motto	*Perstare et praestare* (Latin)
Motto in English	To persevere and to excel
Type	Private[1]
Established	1831[1]
Endowment	$4.1 billion (2018)[2]

View source for New York University

← New York University

You do not have permission to edit this page, for the following reason:

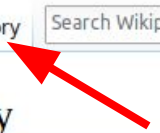
This page is currently semi-protected so that only established registered users can edit it.

B I Advanced Special characters Help

```
}}

'''New York University''' ('''NYU''') is a [[private university|private]] [[nonprofit
institution|nonprofit]] [[research university]] based in [[New York City]]. Founded in 1831,
NYU's primary campus is in [[Greenwich Village]] with other campuses throughout New York
City.<ref name="Schools and Colleges">{{cite web|url=http://www.nyu.edu/academics/schools-and-
colleges.html|title=Schools and Colleges|last=|first=|date=|website=New York
University|publisher=New York University|accessdate=December 30, 2017}}</ref>
<ref>{{Citation|last=New York University|title=Mayor Bloomberg: It's hard to differentiate
where NYU stops and NYC starts|date=August 15, 2013|url=https://www.youtube.com
/watch?v=sBs_t8L64Is|accessdate=January 25, 2017}}</ref> NYU students can also study abroad at
its degree-granting campuses in [[NYU Abu Dhabi]] and [[NYU Shanghai]], as well as its 11
academic centers in [[Accra]], [[Berlin]], [[Buenos Aires]], [[NYU Florence|Florence]], [[NYU
London|London]], [[Madrid]], [[NYU Paris|Paris]], [[Prague]], [[Sydney]], [[Tel Aviv]], and
[[Washington, D.C.]]<ref>{{cite web|url=http://www.nyu.edu/global/global-academic-
centers.html|title=Global Academic Centers|website=New York University}}</ref><ref name="The
Global Network University">{{cite web |url=http://www.nyu.edu/global/the-global-network-
university.html |title=The Global Network University |website=New York University
|publisher=New York University |accessdate=August 30, 2013|archiveurl=https://web.archive.org
/web/20130825101657/http://www.nyu.edu/global/the-global-network-
university.html|archivedate=August 25, 2013}}</ref>

In 2018, NYU was ranked amongst the top 30 universities internationally by the ''[[Academic
Ranking of World Universities]]'', ''[[Times Higher Education World University Rankings]]'',
and ''[[U.S. News & World Report]]''.<ref>https://www.usnews.com/education/best-global-
```



New York University: Revision history

Help

View logs for this page (view filter log)

Show revision history

From year (and earlier): 2018 From month (and earlier): all Tag filter: Show

External tools: Find addition/removal · Find edits by user · Page statistics · Pageviews · Fix dead links

For any version listed below, click on its date to view it.

For more help, see Help:Page history and Help:Edit summary.

(cur) = difference from current version, (prev) = difference from preceding version,

m = minor edit, → = section edit, ← = automatic edit summary

(newest | oldest) View (newer 50 | older 50) (20 | 50 | 100 | 250 | 500)

Compare selected revisions

- (cur | prev) 22:29, 12 November 2018 Tom Stevens (talk | contribs) . . (121,369 bytes) (+428) . . *(→Athletics) (Tag: Visual edit)*
- (cur | prev) 04:58, 10 November 2018 GreenC bot (talk | contribs) . . (120,941 bytes) (-111) . . *(Wayback Medic 2.1)*
- (cur | prev) 07:25, 8 November 2018 Kiwi128 (talk | contribs) . . (121,052 bytes) (-23) . . *(source official website URL from Wikidata)*
- (cur | prev) 04:51, 4 November 2018 ElKevbo (talk | contribs) . . (121,075 bytes) **(+1,085)** . . *(no, we don't need all of this information in the lead; discuss in Talk)*

This November is the Wikipedia Asian Month. Come join us.

Talk:New York University

From Wikipedia, the free encyclopedia

This is the talk page for discussing improvements to the New York University article.

This is not a forum for general discussion of the article's subject.

Semi-protected edit request on 22 June 2018 [edit]

Add after first paragraph of article: "With an undergraduate acceptance rate of 19 percent for the Class of 2022, NYU is becoming one of the most exclusive institutions in the United States." Ssv289 (talk) 16:18, 22 June 2018 (UTC)

Not done: please provide reliable sources that support the change you want to be made.--QueerFilmNerd (talk) 17:38, 22 June 2018 (UTC)

Archives

- Archive 1: Aug 2004–Jan 2007
- Archive 2: Jan 2007–present

Search

This edit request has been answered. Set the `|answered=` or `|ans=` parameter to **no** to reactivate your request.

Public Content

- XML dumps
- Mediawiki API
- SQL Replicas

https://dumps.wikimedia.org/

```
[[Artificial intelligence]] is a recurrent
theme in [[science fiction]], whether
[[utopia]]n, emphasising the potential
benefits, or [[dystopian]], emphasising the
dangers.

The notion of machines with human-like
intelligence dates back at least to [[Samuel
Butler (novelist)|Samuel Butler]]'s 1872
novel ''[[Erewhon]]''.

==Background==
[[File:Didrachm Phaistos obverse
CdM.jpg|thumb|A [[didrachm]] coin depicting
the winged [[Talos]], an [[automaton]] or
[[artificial being]] in [[ancient
Greece|ancient Greek]] myth, c. 300 BC]]
```

WIKIMEDIA
FOUNDATION

Quarry / Sql Replicas

Quarry is a public querying interface for Wiki Replicas, a set of live replica SQL databases of public Wikimedia Wikis.

- https://quarry.wmflabs.org
- https://upload.wikimedia.org/wikipedia/commons/9/94/MediaWiki_1.28.0_database_schema.svg

TED talk links, it:Wikipedia

Fork

It's collaborative and searchable

Fork of TED talk links, it:Wikipedia by Pigsonthewing

This query is marked as a draft by Diego (WMF). Publish

Click to add description...

SQL

```
use itwiki_p;
SELECT page_title, el_to
FROM externallinks
join page on page_id=el_from and page_namespace=0
where el_to like "http://www.ted.com/talks/%" or el_to like
"https://www.ted.com/talks/%" or el_to like "http://ted.com/talks/%" or
el_to like "https://ted.com/talks/%"
```

Resultset (112 rows)

Download data

Show 100 entries

Search:

page_title	el_to
William_Kamkwamba	http://www.ted.com/talks/william_kamkwamba_on_building_a_windmill.html
WikiHouse	http://www.ted.com/talks/alastair_parvin_architecture_for_the_people_by_the_people.html
Vita_personale_di_Leonardo_da_Vinci	http://www.ted.com/talks/siegfried_woldhek_shows_how_he_found_the_true_face_of_leonardo.html
Vilayanur_S._Ramachandran	http://www.ted.com/talks/lang/ita/vilayanur_ramachandran_on_your_mind.html
Vilayanur_S._Ramachandran	http://www.ted.com/talks/lang/ita/vs_ramachandran_the_neurons_that_shaped_civilization.html

Wikidata

Wikidata is a free and open knowledge base that can be read and edited by both humans and machines.

Wikidata acts as central storage for the **structured data** of its Wikimedia sister projects including Wikipedia, Wikivoyage, Wikisource, and others.

In October 2018, there 23 M of edits in Wikidata, compared with around 4 M edits in English Wikipedia

- https://www.wikidata.org/
- https://query.wikidata.org/

WIKIMEDIA
FOUNDATION

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction

Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools

What links here
Related changes
Upload file
Special pages
Permanent link
Page information
Wikidata item
Cite this page

New York University

From Wikipedia, the free encyclopedia

Coordinates: 40°43′48″N 73°59′42″W

This article is about the private institution fou... 1 in 1831. For other and similar uses, see University of New York.
"Nyu" redirects here. For the district in Japan, see Nyū District, Fukui. For the Elfen Lied character, see List of Elfen Lied characters § Nyu.

New York University (**NYU**) is a private nonprofit research university based in New York City. Founded in 1831, NYU's primary campus is in Greenwich Village with other campuses throughout New York City.[12][13] NYU students can also study abroad at its degree-granting campuses in NYU Abu Dhabi and NYU Shanghai, as well as its 11 academic centers in Accra, Berlin, Buenos Aires, Florence, London, Madrid, Paris, Prague, Sydney, Tel Aviv, and Washington, D.C.[14][15]

In 2018, NYU was ranked amongst the top 30 universities internationally by the *Academic Ranking of World Universities*, *Times Higher Education World University Rankings*, and *U.S. News & World Report*.[16][17][18] For the class that matriculated in the fall of 2018, NYU received 75,037 applications for its undergraduate programs; this is more applications than any other private college or university in the United States.[19]

Alumni include heads of state, royalty, eminent scientists, inventors and entrepreneurs, media figures, founders and CEOs of Fortune 500 companies, and astronauts.[20][21][22] As of 2018, 37 Nobel Laureates, 7 Turing Award winners, 5 Fields Medalists, over 30 Academy Award winners, over 30 Pulitzer Prize winners, and hundreds of members of the National Academies of Sciences and United States Congress have been affiliated as faculty or alumni. Globally, NYU is ranked 7th by the Times Higher Education World University Rankings for producing alumni who are millionaires, and 4th by

‹ The template *Infobox university* is being considered for merging. ›

New York University

Latin: *Universitas Neo Eboracensis*

Motto	*Perstare et praestare* (Latin)
Motto in English	To persevere and to excel
Type	Private[1]
Established	1831[1]
Endowment	$4.1 billion (2018)[2]

New York University (Q49210)

private research university in New York, NY, United States edit
NYU | University of the City of New York | University of New York

In more languages Configure

Language	Label	Description	Also known as
English	New York University	private research university in New York, NY, United States	NYU University of the City of Ne... University of New York
Spanish	Universidad de Nueva York	universidad estadounidense	
Traditional Chinese	紐約大學	No description defined	
Chinese	纽约大学	私立研究型大学在美国纽约州纽约市	

All entered languages

Statements

instance of	private university 0 references add reference	edit
	research university 0 references add reference	edit
	private not-for-profit educational institution 1 reference	edit
	add value	

Wikipedia (59 entries) edit

ar جامعة نيويورك
ast Universidá de Nueva York
azb نیویورک یئلیم‌یوردو
az Nyu-York Universiteti
be_x_old Нью-Ёрскі ўнівэрсытэт
be Нью-Ёркскі ўніверсітэт
bg Нюйоркски университет
bn নিউ ইয়র্ক বিশ্ববিদ্যালয়
ca Universitat de Nova York
cs New York University
da New York University
de New York University
eml Università 'd New York
en New York University
eo Universitato de Novjorko
es Universidad de Nueva York
et New Yorgi Ülikool
eu New Yorkeko Unibertsitatea
fa دانشگاه نیویورک
fi New Yorkin yliopisto
fr Université de New York
gl Universidade de Nova York
he אוניברסיטת ניו יורק
hi न्यूयॉर्क विश्वविद्यालय
hu New York Egyetem
hy Նյու Յորքի համալսարան
id Universitas New York

image

NYU07.JPG

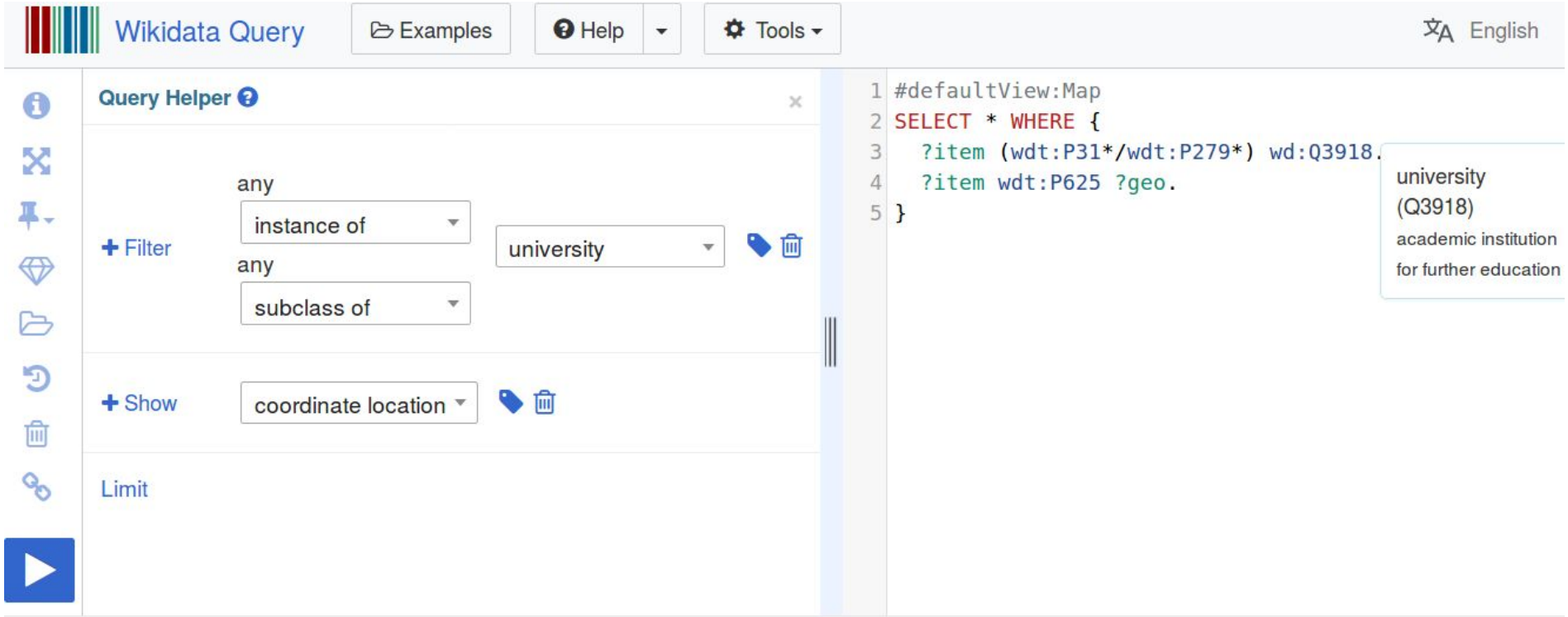
2,592 × 1,952; 2.42 MB

0 references

WIKIMEDIA
FOUNDATION

founded by	Albert Gallatin 0 references
country	United States of America 0 references
located in the administrative territorial entity	Manhattan 1 reference
coordinate location	40°43'48"N, 73°59'42"W 0 references

https://query.wikidata.org



WIKIMEDIA FOUNDATION

Leaflet | Wikimedia | © OpenStreetMap contributors

Wikidata Query | Examples | Help | Tools | English

Query Helper

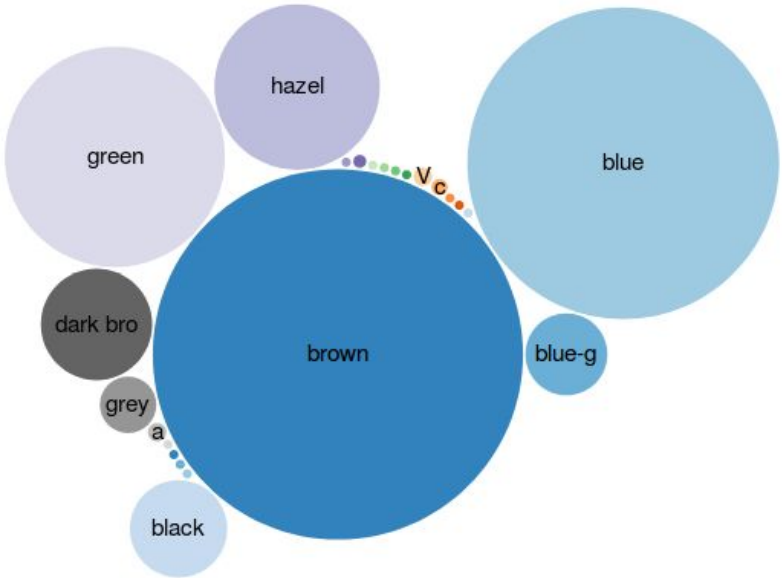
+ Filter | instance of | human

+ Show | eye color

Limit

```
#Popular eye colors among humans
#added before 2016-10
#illustrates bubblechart view, count

#defaultView:BubbleChart
SELECT ?eyeColorLabel (COUNT(?human) AS ?count)
WHERE
{
    ?human wdt:P31 wd:Q5.
    ?human wdt:P1340 ?eyeColor.
    SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAG
}
GROUP BY ?eyeColorLabel
```



Wikimedia Commons

https://commons.wikimedia.org/wiki/Main_Page

Wikimedia Commons (or simply Commons) is an online repository of free-use images, sound, and other media files.

As of 2018, the repository contains over **44 million free** media files, managed and editable by registered volunteers. In July 2013, the number of edits on Commons reached 100,000,000

WIKIMEDIA FOUNDATION

Wikimedia Statistics

Live dashboards about:

- Editing
- Reading
- Users

https://stats.wikimedia.org/v2

WIKIMEDIA
FOUNDATION

Explore topics

Search or Browse questions and pick one to see answers

Wiki

Wikipedia – English

Metrics

Absolute bytes diff

Edited pages

Net bytes difference

Pages to date

Filter and Split

Overall **total**

Split by **editor type**

Anonymous

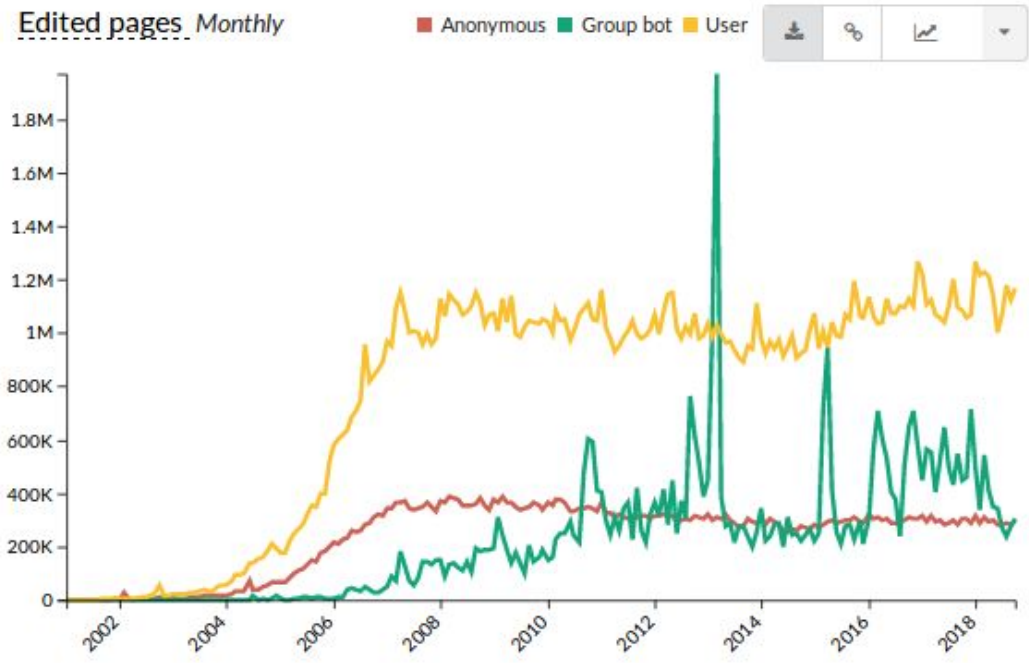
Group bot

Name bot

User

Split by **page type**

Split by **activity level**



Average: 1.22M ↑ 3063484.48% over this time range.

The number of pages edited, excluding redirect pages. More info about this metric.
You can see or add annotations for this data on this wiki article.

Page Views

This page documents the Pageview API (v1), a public API developed and maintained by the Wikimedia Foundation that serves analytical data about article pageviews of Wikipedia and its sister projects. With it, you can get pageview trends on specific articles or projects; filter by agent type or access method, among other things.

- https://wikitech.wikimedia.org/wiki/Analytics/AQS/Pageviews

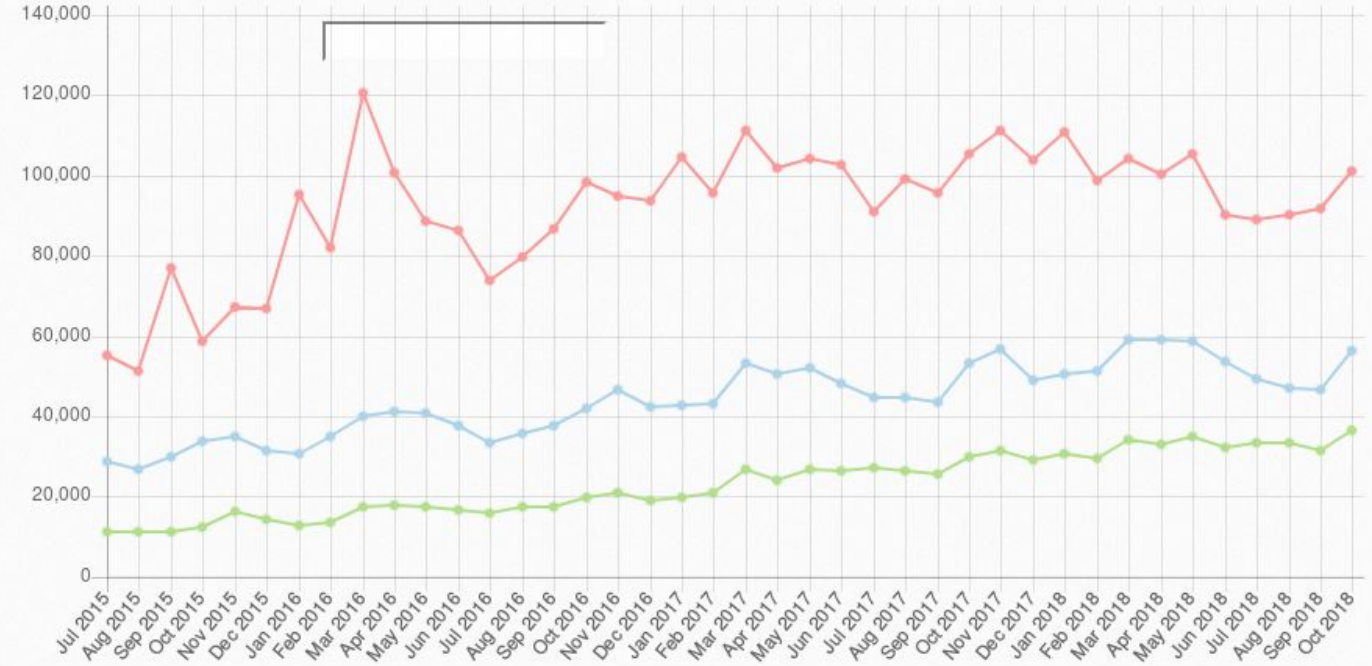
Pageviews Analysis

Comparison of pageviews across multiple pages

Pages Enter up to 10 pages × Clear

× Random_forest × Gradient_boosting × Deep learning

Chart type | Permalink | Download | Show values | Begin at zero | Logarithmic scale



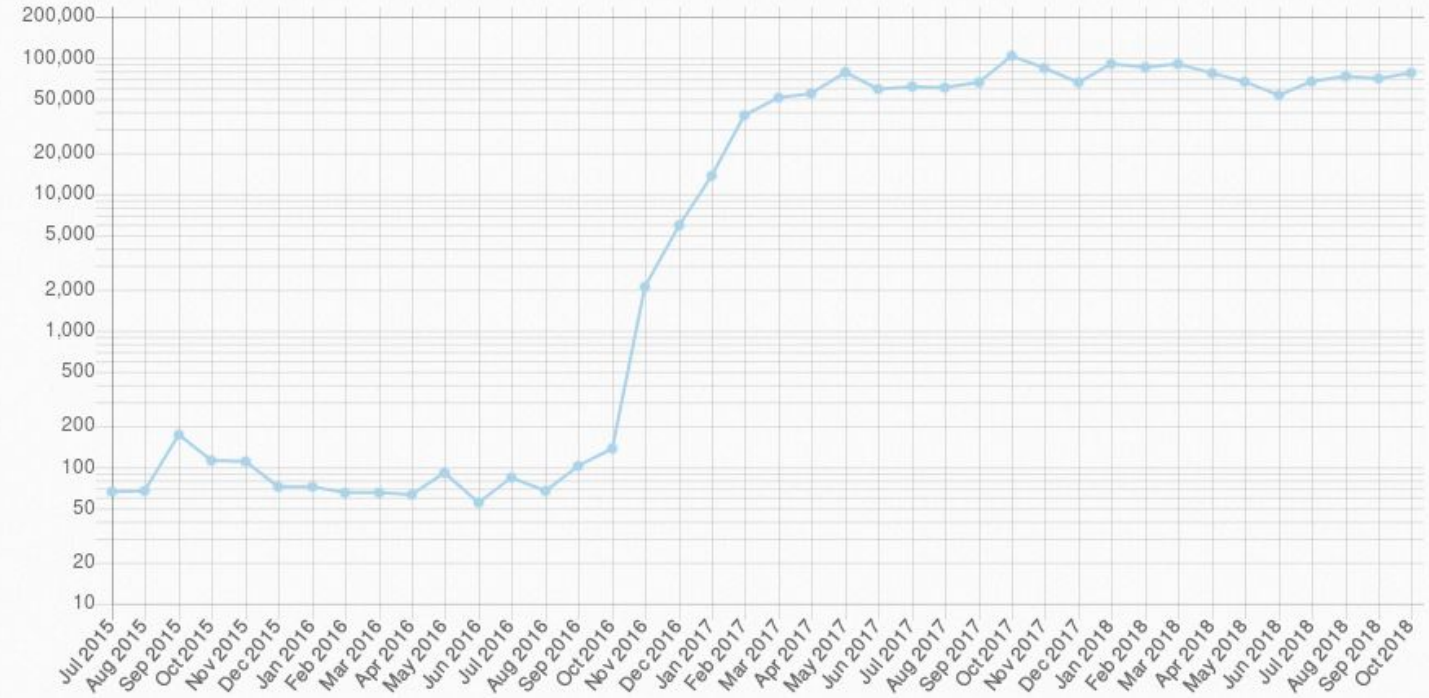
Pageviews Analysis

Comparison of pageviews across multiple pages

Pages Enter up to 10 pages × Clear

× Fake news

Chart type | Permalink | Download ▾ | Show values | Begin at zero | Logarithmic scale



Topviews Analysis

Most viewed pages of a project

Date type: Monthly | **Date**: July 2018 | **Project**: en.wikipedia.org | **Platform**: All

Excluded pages: hover-to-exclude

☐ Show mobile percentages ☑ Show only mainspace pages

Search

Some pages listed may be false positives. known-false-positives-text

Permalink | Download

Rank	Page	Edits	Editors	Pageviews
1.	2018 FIFA World Cup	533	176	8,614,841
2.	Exo (band)	70	14	5,786,931
3.	Croatia	74	28	4,961,832
4.	Kylian Mbappé	224	53	3,621,830
5.	FIFA World Cup	178	74	3,591,619
6.	List of FIFA World Cup finals	124	66	3,522,853
7.	Cristiano Ronaldo	345	98	3,436,411
8.	Deaths in 2018	1,986	197	3,297,083
9.	Demi Lovato	40	20	2,943,525
10.	Ant-Man and the Wasp	625	152	2,916,170

Langviews Analysis

Pageviews of a page across all languages

Do another query

Jair Bolsonaro 10/22/2018 - 11/11/2018

List | Chart

Permalink | Download

#	Language	Page title	Badges	Pageviews	Daily average
totals	**num-languages**	**unique-titles**		**2,388,527**	**113,739**
1	pt	Jair Bolsonaro		1,156,777	55,085
2	en	Jair Bolsonaro		583,581	27,790
3	es	Jair Bolsonaro		217,074	10,337
4	fr	Jair Bolsonaro		142,608	6,791
5	de	Jair Bolsonaro		89,745	4,274

https://stats.wikimedia.org/v2/#/es.wikipedia.org/reading/pageviews-by-country

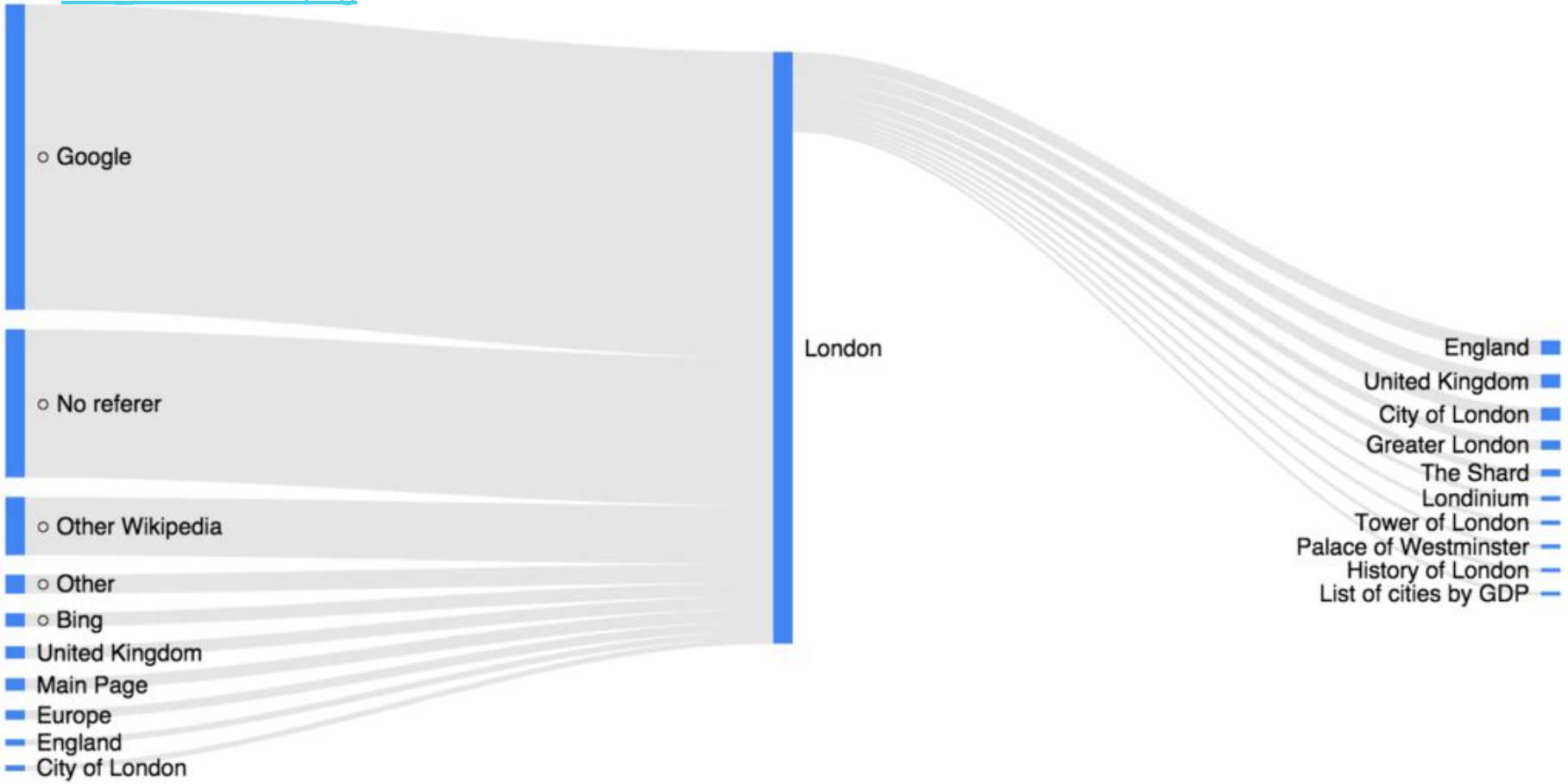
Click dataset

The Wikipedia Clickstream dataset contains counts of (referer, resource) pairs extracted from the **request logs of Wikipedia**. A referer is an HTTP header field that identifies the address of the webpage that linked to the resource being requested. The data shows how people get to a Wikipedia article and what links they click on.

Automatic monthly releases of datasets for Wikipedia in **English, Russian, German, Spanish and Japanese.**

- https://meta.wikimedia.org/wiki/Research:Wikipedia_clickstream

https://upload.wikimedia.org/wikipedia/commons/thumb/0/02/London_clickstream.png/1024px-London_clickstream.png



Ores

ORES (Objective Revision Evaluation Service) is a web service and API that provides machine learning as a service for Wikimedia projects maintained by the Scoring Platform team.

- http://ores.wikimedia.org

WIKIMEDIA
FOUNDATION

Ores

http://ores.wmflabs.org/scores/enwiki/damaging/638307884

```
"damaging": {
"prediction": false,
"probability": {
"false": 0.8683,
"true": 0.1317
}
}
```

English Wikipedia revisions as seen by editors (left) and as scored by ORES' "damage" prediction model (right).

https://blog.wikimedia.org/2015/11/30/artificial-intelligence-x-ray-specs/

WIKIMEDIA FOUNDATION

Wikimedia Toolforge

Toolforge is a hosting environment for developers working on services that provide value to the Wikimedia movement. These services allow developers to easily do ad hoc analytics, administer bots, run webservices, and generally create tools to help editors and other volunteers in their work. The environment also includes access to a variety of data services.

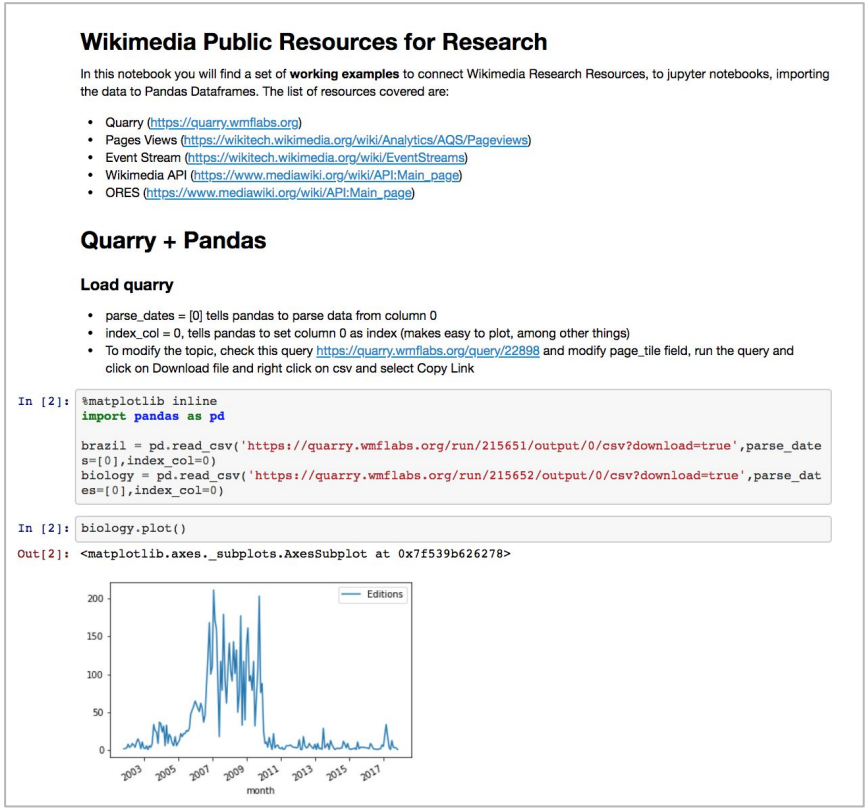
- https://tools.wmflabs.org/

WIKIMEDIA
FOUNDATION

PAWS

https://paws.wmflabs.org

A Web Shell (PAWS) is a Jupyter notebooks deployment that has been customized to make interacting with Wikimedia wikis easier.



Wikimedia Public Resources for Research

In this notebook you will find a set of **working examples** to connect Wikimedia Research Resources, to jupyter notebooks, importing the data to Pandas Dataframes. The list of resources covered are:

- Quarry (https://quarry.wmflabs.org)
- Pages Views (https://wikitech.wikimedia.org/wiki/Analytics/AQS/Pageviews)
- Event Stream (https://wikitech.wikimedia.org/wiki/EventStreams)
- Wikimedia API (https://www.mediawiki.org/wiki/API:Main_page)
- ORES (https://www.mediawiki.org/wiki/API:Main_page)

Quarry + Pandas

Load quarry

- parse_dates = [0] tells pandas to parse data from column 0
- index_col = 0, tells pandas to set column 0 as index (makes easy to plot, among other things)
- To modify the topic, check this query https://quarry.wmflabs.org/query/22898 and modify page_tile field, run the query and click on Download file and right click on csv and select Copy Link

In [2]:
```
%matplotlib inline
import pandas as pd

brazil = pd.read_csv('https://quarry.wmflabs.org/run/215651/output/0/csv?download=true',parse_date
s=[0],index_col=0)
biology = pd.read_csv('https://quarry.wmflabs.org/run/215652/output/0/csv?download=true',parse_dat
es=[0],index_col=0)
```

In [2]:
```
biology.plot()
```

Out[2]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f539b626278>
```


WIKIMEDIA FOUNDATION

Event Stream

EventStreams is a web service that exposes continuous streams of **structured event data.** It does so over HTTP using chunked transfer encoding following the Server-Sent Events protocol. EventStreams can be consumed directly via HTTP, but is more commonly used via a client library.

- https://wikitech.wikimedia.org/wiki/EventStreams
- http://listen.hatnote.com/

WIKIMEDIA
FOUNDATION

Working Examples

http://paws-public.wmflabs.org/paws-public/User:Diego_(WMF)/WikiMediaPublicTools.ipynb

https://github.com/digitalTranshumant/Wiki-examples

WIKIMEDIA
FOUNDATION