

Undocumented APIs

How to get the data behind a webpage or map
without scraping

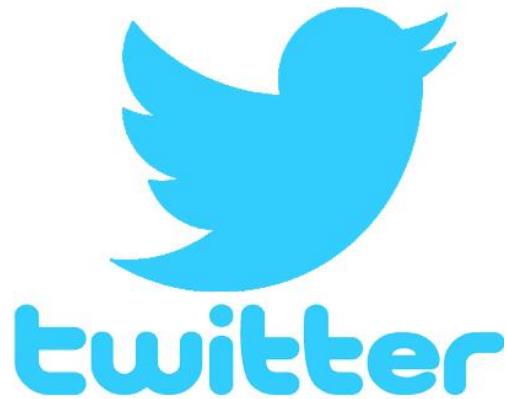
By the end of this you should know...

- **What APIs** are — and what an ‘undocumented’ one is
- **How** they’re used to find stories
- **Basic techniques** for finding undocumented APIs
- **How to get the data out**

1. What is an API?

Application Programming Interface

- Application (for apps, e.g. Twitter, websites)
- Programming (...using code...)
- Interface (...to interact)



TweetDeck

“Give me tweets
with the hashtag
#eurovision & in
descending order”



“Here are tweets
with the hashtag
#eurovision & in
order=descending”

(But that's not why we're interested in them.)

An API is... a source of information

- **Data is often made available through querying an API**
- More **flexible** than spreadsheets: different data and different shapes
- More than one dataset: **multiple responses** to multiple queries
- More **efficient** than scraping: no parsing of HTML
- More **reliable**: HTML likely to be updated more often than API
- **'Live'** data: can be used to feed dashboards and interactives

open corporates

“For this keyword,
give me a list of
companies & their
directors”



JADE: MYANMAR'S "BIG STATE SECRET"

The biggest natural resources heist in modern history?

"Not only was it critical that OpenCorporates had this information (much of the information was removed from the official register during the investigation), but the fact that it was machine-readable data, available via an API (data service), and programmatically combinable with other data was essential to discover the hidden connections between the key actors and the jade industry. Global Witness was able to analyse this data with the help of Open Knowledge."

Rail season tickets cost 10% of net pay

By Daniel Wainwright

BBC News

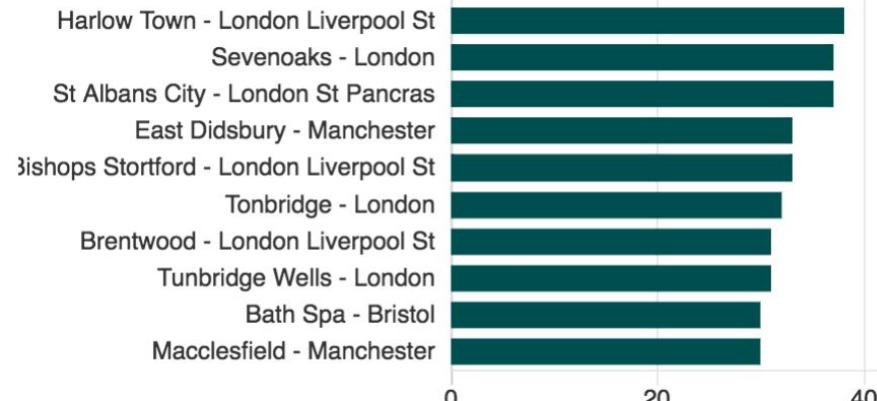
12 September 2016



Most expensive rail season tickets

Prices of annual passes mile mile travelled for full time workers

■ Pence per mile



Source: campaign for Better Transport/Google Places

BBC

Festivals dominated by male acts, study shows, as Glastonbury begins

By Pete Sherlock and Paul Bradshaw
BBC News

⌚ 22 June 2017



Spotify Web API

API documentation tells you how to use it

Postcodes.io Documentation About Explore

Documentation

Overview
Postcodes.io is a free Postcode lookup API and geocoder for the UK

Endpoint
All services can be accessed from the following HTTP[S] endpoint
<https://api.postcodes.io>



DOCUMENTACIÓN DE BESOCCKER API

Ponemos a tu disposición una amplia documentación técnica con toda la información y las herramientas necesarias, para que puedas probar toda la API de BeSoccer de manera ágil y sencilla.

[Ver documentación aquí](#)

But some APIs aren't public

- **Undocumented APIs** ‘feed’ websites data
- An **internal database**, without public documentation
- E.g. search query results, map points, chart data, Amazon's autosuggest, etc.
- Its public use makes it possible to access its data

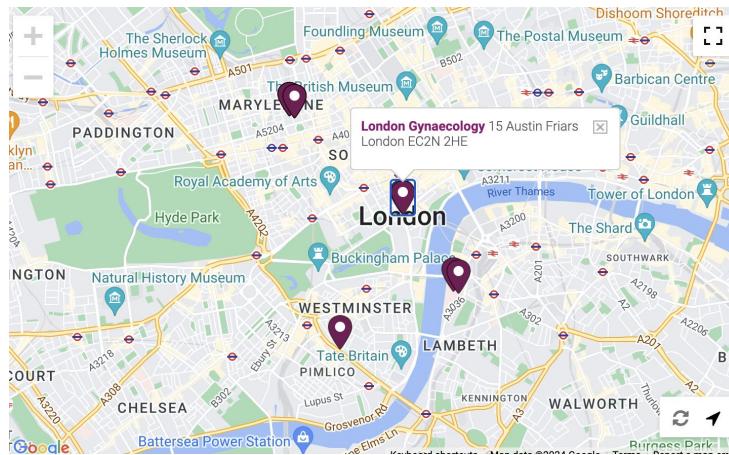
Postcode lottery of menopause care revealed



Menopause Support founder Diane Danzebrink said women were being pushed to seek private care

By Sam Ferguson and Lauren Woodhead

BBC Shared Data Unit



BBC analysis of British Menopause Society data showed 59% of Clinical Commissioning Group areas (CCGs) in England had no NHS clinics, including all of Devon and Cornwall and large areas in the North.

The data also showed almost a third of existing NHS clinics were either in London or Surrey.

<https://www.bbc.co.uk/news/uk-england-62849438>

https://github.com/BBC-Data-Unit/Hormone_replacement_therapy

<https://thebms.org.uk/find-a-menopause-specialist/>



Google the Giant

Google Has a Secret Blocklist that Hides YouTube Hate Videos from Advertisers—But It's Full of Holes

Many well-known White supremacist and White nationalist terms and slogans were not blocked

By Leon Yin and Aaron Sankin

Most hate terms on our list were not blocked

link

Google Ads responses for hate-related terms

All hate-related terms (86)



Ring's Hidden Data Let Us Map Amazon's Sprawling Home Surveillance Network

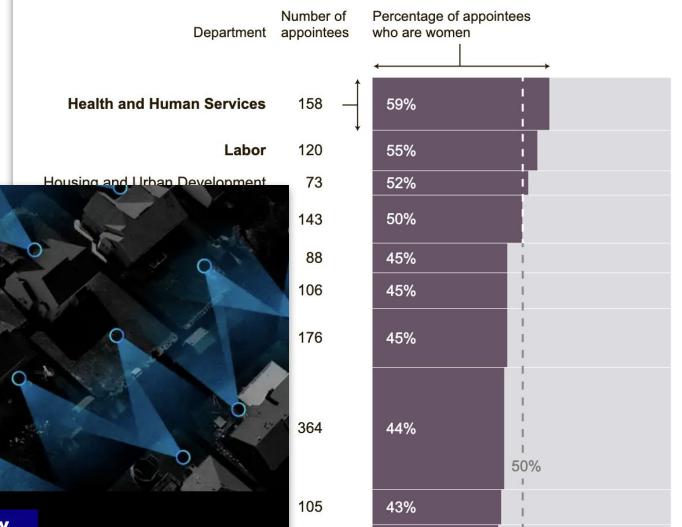
Gizmodo examined more than 20,000 Ring cameras to find a sprawling surveillance net covering the nation.

Published: January 8, 2013

Women in the Obama Administration

While the White House itself employs about an equal number of men and women under President Obama, the gender ratio of appointees in 11 of the 15 cabinet departments favors men. [Related Article »](#)

Departments in **bold** have a female secretary or acting secretary.



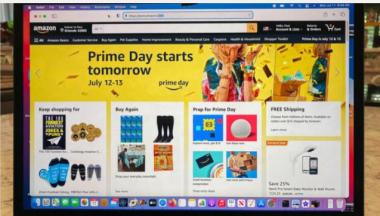
"Gizmodo located cameras in unnerving proximity to such sensitive buildings, including a clinic offering abortion services and a legal office that handles immigration and refugee cases."

[The Markup: Investigating YouTube blocklists](#)

[Gizmodo: checking Ring's privacy safeguards](#)

[NYT: Women in the Obama Administration \(method\)](#)

Amazon's Advantage



Amazon's Advantage

Amazon Ranks Its Own Products First, FTC Lawsuit Says

Antitrust complaint comes two years after a Markup investigation reached the same conclusion about Amazon search results

September 28, 2023 08:00 ET



Amazon's Advantage

The Markup Wins Loeb Award for Amazon's Advantage

Our series revealed that Amazon placed its own in-house and exclusive products ahead of third-party items with more reviews and higher ratings

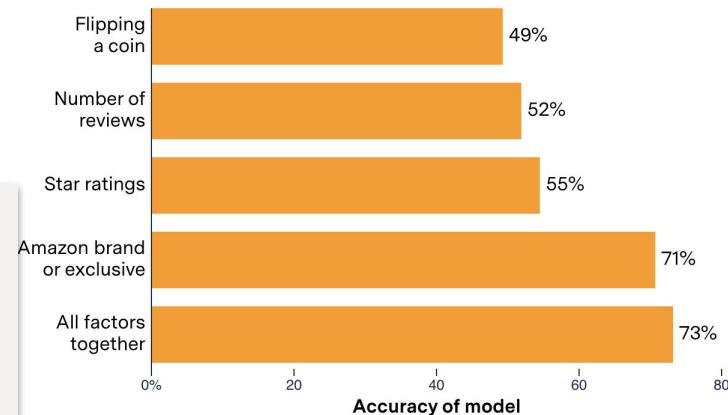
Amazon's Advantage

Amazon Puts Its Own “Brands” First Above Better-Rated Products

The online giant gives a leg up to hundreds of house brand and exclusive products that most people don't know are connected to Amazon

Predicting the top spot

Being an Amazon brand or exclusive influenced ranking more than star ratings or reviews.



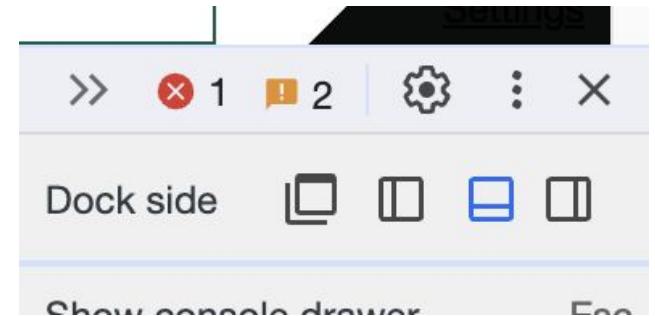
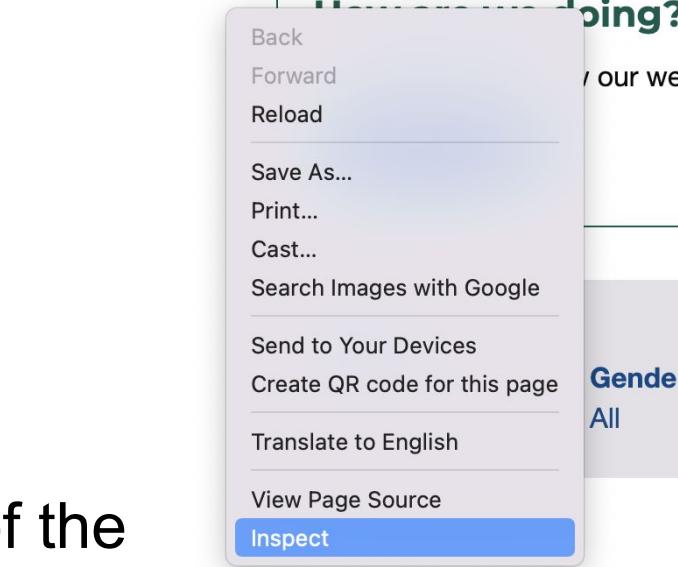
We used machine learning to try to predict which product Amazon put first in search results based on various factors. Source: The Markup/Amazon.com

2. Finding an undocumented API

Note:
You must be using Firefox (best) or Chrome

Opening the Inspector

1. **Right-click** on a webpage — select **Inspect**
2. The **Inspector** will open across half of the page
3. You can change which side it is ‘docked’ by clicking on the three dots in the upper left corner



Use the Network tab

The Inspector has several tabs:

Elements; Console; Sources; Network; Performance...

The **Network** tab is the one that we want: it can show us any files that the webpage is loading



Reload the page with the Network tab open

- The Network tab only **records** while it's **open**
- When you **reload the page**, you will see all the files it uses: images, video, scripts, cookies, etc.
- **Use the sub-tabs to filter** to different types of file
- Look in the **Fetch/XHR** tab for API requests (XHR in Firefox)



Case study: Fatal Incident reports

Fatal Incident reports

This is a searchable record of our final reports, published on our website once they have been shared with the family of the deceased person, and the coroner's inquest has taken place. Names have been removed to protect the privacy of the individuals concerned. For deaths after 1 March 2015, the name of the deceased remains in the report, but other names have been anonymised.

If you would like an update on a report that you cannot find or if you spot any errors, please email: ppocomms@ppo.gov.uk. Please use the filters to limit the display to specific case types and the sort buttons to reorder by either date of death or date it went on the website.

[Hide details](#)

St Leonards

Published: 07/04/2021

Approved premises

Date of death: 25/07/2020
Cause: Self-inflicted
Gender: Male
Age: 22-30

[PPO Report](#) (PDF, 600 KB)

[Action Plan](#) (PDF, 135 KB)

Bure

Published: 07/06/2021

Prison

Date of death: 17/05/2020
Cause: Self-inflicted
Gender: Male
Age: 41-50

[PPO Report](#) (PDF, 570 KB)

[Action Plan](#) (PDF, 129 KB)

Durham

Published: 30/07/2021

Prison

Date of death: 02/05/2020
Cause: Self-inflicted
Gender: Male
Age: 41-50

[PPO Report](#) (PDF, 636 KB)

[Action Plan](#) (PDF, 147 KB)

Go to tinyurl.com/3mywvs3w
or ppo.gov.uk/document/fii-report/

An ‘unscrapable’ page...

- If you write a normal scraper for this page, **it won’t work**
- Why? Because the page loaded by the scraper is just the ‘skeleton’ HTML **before external data is loaded**
- **Check** by adding a line to print the whole HTML of the page, and see if the data is there
- So we check the **Inspector**...

Open the Network tab > Fetch/XHR (+ reload)

The screenshot shows the Network tab in the Chrome DevTools interface. A red box highlights the 'Network' tab in the top navigation bar. Another red box highlights the 'XHR' tab under the filter section. A third red box highlights the table below, which lists network requests. The table has columns: Name, Status, Type, Initiator, Size, Time, and Waterfall. One row is visible, showing a request to 'admin-ajax.php?' with a status of 200, type 'xhr', initiator 'jquery...', size 4.2..., and time 2.1... ms.

Name	Status	Type	Initiator	Size	Time	Waterfall
admin-ajax.php?...	200	xhr	jquery....	4.2...	2.1...	

Try: right-click > Open in new tab

The screenshot shows the Chrome DevTools Network tab. At the top, there are tabs for Elements, Console, Sources, Network (which is selected), and Performance. Below the tabs are various controls like a red circle, a search icon, and a filter input. The main area displays a timeline from 500 ms to 2500 ms. A context menu is open over a network entry for 'admin-ajax.php?...'. The menu items are: 'Open in Sources panel' (grayed out), 'Open in new tab' (highlighted in blue), and 'Clear browser cache'.

Name	Sta...	Type	...
admin-ajax.php?...	200	xhr	j

[Example linked file](#)

What is it? A HTML file with the data

Coster, Stephen

Lewes

Prison

Published: 5 January 2024

Date of death: 5 May 2022

Date of inquest: 4 Dec 2023

Cause: Natural causes

Gender: Male

Age: 41-50

- [PPO Report](#) (PDF, 355 KB)
- [Action Plan](#) (PDF, 121 KB)

Baker, Brian

Sudbury

Prison

Published: 5 January 2024

Date of death: 1 May 2017

Date of inquest: 13 Dec 2023

Cause: Natural causes

Gender: Male

Age: 61+

- [PPO Report](#) (PDF, 419 KB)
- [Action Plan](#) (PDF, 13 KB)

Coster, Stephen

Lewes

Prison

Published: 5 January 2024

Date of death: 5 May 2022

Date of inquest: 4 Dec 2023

Cause: Natural causes

Gender: Male

Age: 41-50

[PPO Report](#) (PDF, 355 KB)

[Action Plan](#) (PDF, 121 KB)

Baker, Brian

Sudbury

Prison

Published: 5 January 2024

Date of death: 1 May 2017

Date of inquest: 13 Dec 2023

Cause: Natural causes

Gender: Male

Age: 61+

[PPO Report](#) (PDF, 419 KB)

[Action Plan](#) (PDF, 13 KB)

Look at the URL (more later)

https://poo.gov.uk/wp-admin/admin-ajax.php?
action=update_tiles&
queryParams=%7B%22document_type%22%3A%22fii-report
%22%2C%22posts_per_page%22%3A50%2C%22paged%2
2%3A1%2C%22order%22%3A%22DESC%22%2C%22order
by%22%3A%22meta_value%22%2C%22meta_key%22%3A
%22document-date%22%7D
&_=1704708818980

Equals signs indicate parameters (**ingredients**)

The question mark indicates a **query**

Each ingredient is separated by an ampersand: &

Parts beginning with % are ways of expressing characters like “ in a URL:
%22 is “, %7B is {, %2C is comma

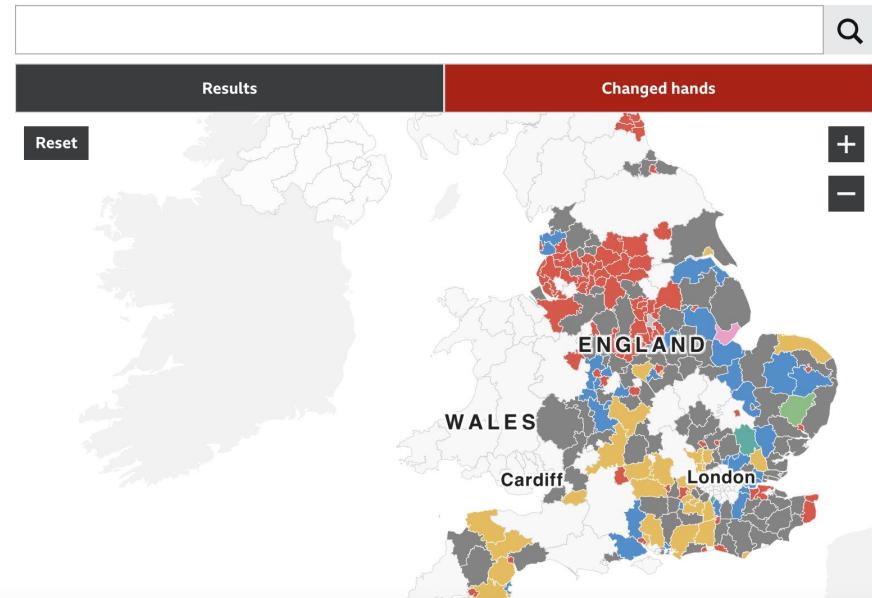
Case study: Data behind a map

tinyurl.com/2swenkpd

bbc.co.uk/news/election/2023/england/results

Who won in my area?

Enter your postcode, or the name of your council area to find out



Open the Network tab > Fetch/XHR (+ reload)

Name	Status	Type
consent-banner	200	xhr
config	200	xhr
Au22auYqZXzjcyGBqtzfa.json	200	xhr
userinfo	200	xhr
key.min.json	200	fetch
locationByGssid.json	200	fetch
council_control_history.min.json	200	fetch
party_hex.min.json	200	fetch

How do we know which one we need?

- **We don't.** It's a process of trial and error.

However, look for:

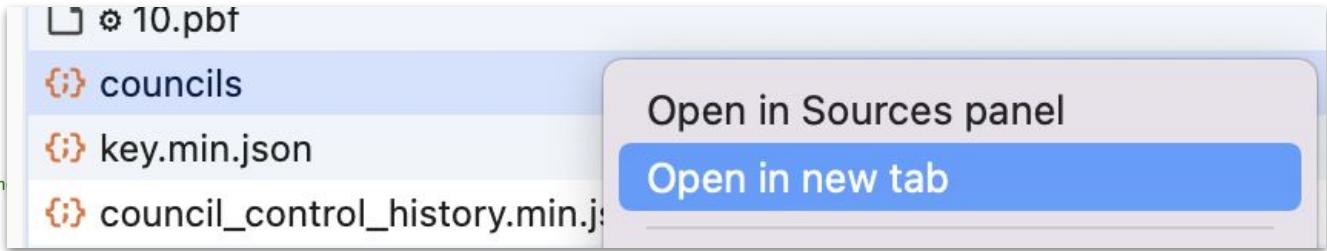
- Files which are large (you can **sort by size**)
- Files with promising **names**
- Files with **?** in (indicates a query to an API)
- Files ending in **.json** or **.xml** or **.csv**

Sorted by size: names that jump out

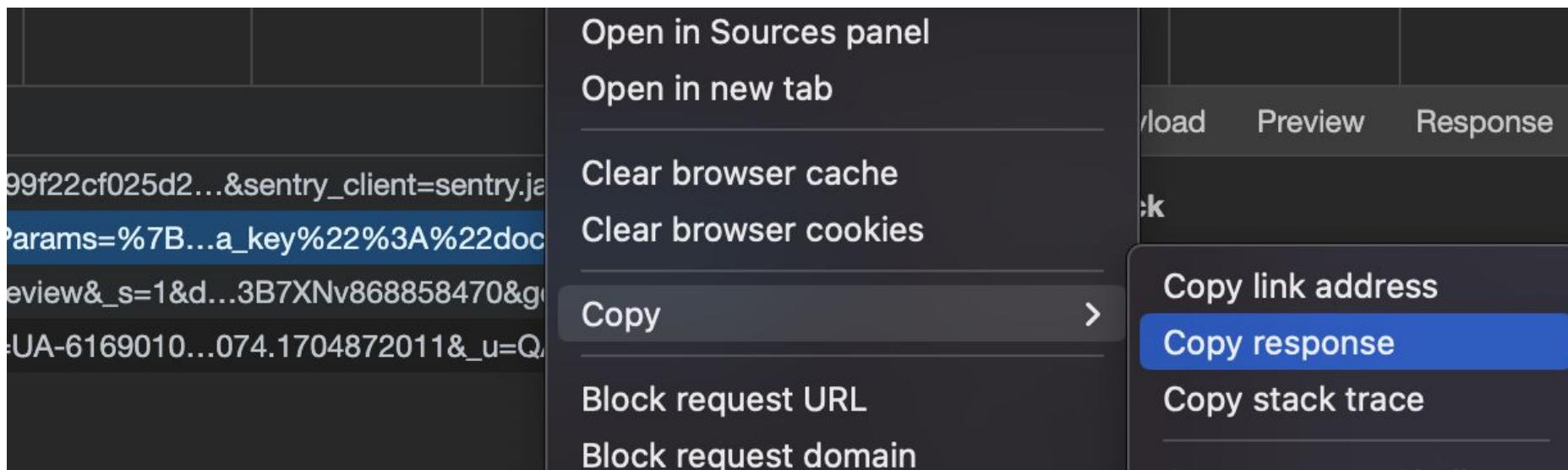
Name	Status	Type	Initiator	Size	Tir
Do not clear log on page reload / navigation					
consent-banner	200	xhr	bundle-defaultVendors.e2e...	6.1 kB	
config	200	xhr	bundle-defaultVendors.e2e...	4.2 kB	
Au22auYqZXzjcyGBqtzfa.json	200	xhr	bundle-pwa.d4ced85....js:1	505 B	
councils	200	fetch	common.js:57	491 B	
results	200	fetch	common.js:57	491 B	
uk.topo.json?v=1	200	fetch	common.js:57	470 B	
kev.min.json	200	fetch	common.js:57	470 B	
council_control_history.min.json	200	fetch	common.js:57	470 B	

Right-click > Open in new tab > Save as

```
,  
  groups: [  
    {  
      id: "letter-a",  
      heading: "A",  
      cards: [  
        {  
          title: "Amber Valley",  
          href: "/news/election/2023/en  
          winnerFlash: {  
            displayMode: "party",  
            year: 2023,  
            prevYear: "previous",  
            newColour: "#E91D0E",  
            prevColour: "#0575C9",  
            flash: "LABOUR GAIN FROM CONSERVATIVE",  
            longFlash: "2023 Labour gain from Conservative",  
            textColour: "#FFFFFF",  
            partyName: "Labour",  
            flashBold: "LABOUR GAIN",  
            flashRegular: " FROM CONSERVATIVE",  
            winnerPartyCode: "LAB",  
            prevWinnerPartyCode: "CON"  
          },  
          context: null  
        },  
        {  
          title: "Arun",  
          href: "/news/election/2023/england/councils/E07000224",  
          winnerFlash: {  
            displayMode: "party",  
            year: 2023,  
            prevYear: "previous",  
            newColour: "#646464",  
            prevColour: "#646464",  
            flash: "NO PARTY MAJORITY",  
            longFlash: "2023 no overall control, no change from previous year",  
            textColour: "#FFFFFF",  
            ...  
          }  
        }  
      ]  
    }  
  ]  
}
```



**Or: right-click > copy > copy response
Paste into a text editor and save as .json**



[Example linked file](#)

Chrome tip: install JSONView or JSON Viewer



This will ‘prettyfy’ the JSON so that it’s easier to understand

Some coding knowledge helps...

- **Curly** brackets indicate a **dictionary**: { } - a way of storing data
- A colon indicates a **key-value pair** (keys are like column headings, e.g. "**year**": **2023** means the value 2023 is in the 'year' column)
- **Square** brackets indicate a **list**: []
- Dictionaries can contain lists, and vice versa

For example:

```
"groups": [
  {
    "id": "letter-a",
    "heading": "A",
    "cards": [
      {
        "title": "Amber Valley",
        "href": "/news/election/2023/england/councils/E
07000032",
```

<https://static.files.bbci.co.uk/elections/archive/news/election/2023/england/councils>

3. Getting the data out

Save the file and use a JSON>CSV tool

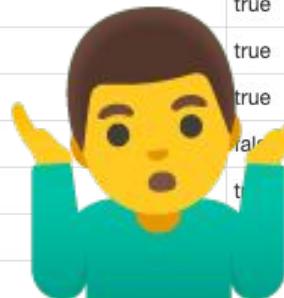
- Search for **JSON to CSV converter**
- Try different tools until you're happy with the results

Results from convertcsv.com/json-to-csv.htm:

Save your result: councils.csv Download Result EOL: CRLF

id	name	enabled
letter-a	A	true
letter-b	B	true
letter-c	C	true
letter-d	D	true
letter-e	E	true
letter-f	F	true
letter-g	G	true
letter-h	H	true
letter-i	I	true
letter-j	J	true
letter-k	K	true
letter-l	L	true

```
links: null,  
key: [  
  {  
    id: "letter-a",  
    name: "A",  
    enabled: true  
  },  
  {  
    id: "letter-b",  
    name: "B",  
    enabled: true  
  },  
  {  
    id: "letter-c",  
    name: "C",  
    enabled: true  
  },  
  {  
    id: "letter-d",  
    name: "D",  
    enabled: true  
  },  
  {  
    id: "letter-e",  
    name: "E",  
    enabled: true  
  },  
  {  
    id: "letter-f",  
    name: "F",  
    enabled: true  
  },  
  {  
    id: "letter-g",  
    name: "G",  
    enabled: true  
  },  
  {  
    id: "letter-h",  
    name: "H",  
    enabled: true  
  },  
  {  
    id: "letter-i",  
    name: "I",  
    enabled: true  
  },  
  {  
    id: "letter-j",  
    name: "J",  
    enabled: true  
  },  
  {  
    id: "letter-k",  
    name: "K",  
    enabled: true  
  },  
  {  
    id: "letter-l",  
    name: "L",  
    enabled: true  
  }]
```



The tool fetches a [list] inside **one branch: key**
Each {} is treated as a row

Results from [data.page/json/csv](#):

R	S	T	U	V	W	X	Y
key_id	key_name	key_enabled	groups_id	groups_heading	groups_cards	groups_cards_title	groups_cards_href
letter-a	A	TRUE	letter-a	A	Amber Valley	/news/electi	party
letter-b	B	TRUE		Arun	/news/electi	party	
letter-c	C	TRUE		Ashfield	/news/electi	party	
letter-d	D	TRUE		Ashford	/news/electi	party	
letter-e	E	TRUE	letter-b	B	Babergh	/news/electi	party
letter-f	F	TRUE		Barnsley	/news/electi	party	
letter-g	G	TRUE		Basildon	/news/electi	party	
letter-h	H	TRUE		Basingstoke	/news/electi	party	
letter-i	I	TRUE		Bassetlaw	/news/electi	party	
letter-j	J	FALSE		Bath & North	/news/electi	party	
letter-k	K	TRUE		Bedford	/news/electi	party	
letter-l	L	TRUE		Blaby	/news/electi	party	
letter-m	M	TRUE		Blackburn with	/news/electi	party	
letter-n	N	TRUE		Blackpool	/news/electi	party	
letter-o	O	TRUE		Bolsover	/news/electi	party	
letter-p	P	TRUE		Bolton	/news/electi	party	
letter-q	Q	FALSE		Boston	/news/electi	party	
letter-r	R	TRUE		Bournemouth	/news/electi	party	
letter-s	S	TRUE		Bracknell Forest	/news/electi	party	
letter-t	T	TRUE		Bradford	/news/electi	party	
letter-u	U	TRUE		Braintree	/news/electi	party	
letter-v	V	TRUE		Breckland	/news/electi	party	
letter-w	W	TRUE		Brentwood	/news/electi	party	

```

key: [
  {
    id: "letter-a",
    name: "A",
    enabled: true
  },
  {
    id: "letter-b",
    name: "B",
    enabled: true
  }
],
groups: [
  {
    id: "letter-a",
    heading: "A",
    cards: [
      {
        title: "Amber Valley",
        href: "/news/election/2023/england/councils/E07000032",
        winnerFlash: {
          displayMode: "party",
          year: 2023,
          prevYear: "previous",
          newColour: "#E91D0E",
          prevColour: "#0575C9",
          flash: "LABOUR GAIN FROM CONSERVATIVE",
          longFlash: "2023 Labour gain from Conservative",
          textColour: "#FFFFFF",
          partyName: "Labour",
          flashBold: "LABOUR GAIN",
          flashRegular: " FROM CONSERVATIVE",
          winnerPartyCode: "LAB",
          prevWinnerPartyCode: "CON"
        },
        context: null
      },
      {
        title: "Arun",
        href: "/news/election/2023/england/councils/E07000224",
        winnerFlash: {
          displayMode: "party",
          year: 2023,
        }
      }
    ]
  }
]

```

Branches are '**flattened**':
groups_cards_title = **groups > cards > title**

Digging deeper

Try this URL: **tinyurl.com/z7cy775a**

Or:

openrent.co.uk/properties-to-rent/bristol-bristol?term=Bristol,%20Bristol

Open **Inspector > Network, XHR, and reload**

Click on a result - get more details

The screenshot shows the Chrome DevTools Network tab with several requests listed. A red box highlights the first request, which is a GET request to `www.openrent.co.uk/search/propertiesbyid?ids=2052134&ids=2065557&ids=2068322&ids=2080130&ids=2016124&ids=2040971&ids=2078771&ids=1584250&ids=2020480&ids=1988999&ids=1989526&ids=2020400&ids=2048034&ids=2063918&ids=2079164&ids=2040033&ids=2056161&ids=1989506&ids=2028329&ids=2017761`. An arrow points from this request to the Headers panel, which is also highlighted with a red box. The Headers panel displays the following details:

Status	200 ⓘ
Version	HTTP/2
Transferred	6.53 kB (10.95 kB size)
Referrer Policy	strict-origin-when-cross-origin

What information can we get?

- **Headers**: the URL of the ‘request’, method etc.
- **Payload** (Chrome): parameters (ingredients) of the request
- **Preview**: of the response...
- **Response**: what was received, i.e. **the data**
- Initiator, Timing, Cookies: not important

Headers: makes parameters clearer (Firefox)

The screenshot shows the Firefox Developer Tools Network panel. The top navigation bar includes tabs for SS, JS, XHR (which is selected), Fonts, Images, Media, WS, and Other. There are also options to Disable Cache and No Throttling. Below the tabs, there are buttons for Headers, Cookies, Request, Response, Timings, Stack Trace, and Security. The Headers tab is currently active, indicated by a blue underline. A 'Filter Headers' input field is present. To the right of the tabs, there are 'Block' and 'Reset' buttons. The main content area displays a list of network requests. The first request listed is a GET request to `https://www.openrent.co.uk/search/propertiesbyid?ids=2047086&ids=2052134&ids=1978919&ids=1978933&ids=2065557&ids=2068322&ids=2080130&ids=2016124&ids=2040971&ids=2078771&ids=1584250&ids=2020480&ids=1988999&ids=1989526&ids=2020400&ids=2048034&ids=2063918&ids=2079164&ids=2040033&ids=2056161`. At the bottom, status information is shown: Status 200 (with a question mark icon) and Version HTTP/2.

Parameters: ingredients in an API query

`https://www.openrent.co.uk/search/propertiesbyid?
ids=2052134
&
ids=2065557
&
ids=2068322&ids=2080130&ids=2016124&
ids=2040971&ids=2078771&ids=1584250&
ids=2020480&ids=1988999&ids=1989526&
ids=2020400&ids=2048034&ids=2063918&ids=207
9164&ids=2040033&ids=2056161&ids=1989506&i
ds=2028329&ids=2017761`

The ingredients are a series of 'ids' that the website wants data on. The value of this one is 2052134

The word before the ? indicates the type of query being asked (the 'endpoint', 'function' or 'method')

Because there's more than one parameter, each is separated by an &

Try editing the query ingredients

[https://www.openrent.co.uk/search/propertiesbyid?](https://www.openrent.co.uk/search/propertiesbyid?ids=2052134)

ids=2052134



What if we delete it all
apart from just one?

Try adapting the URL ingredients

[https://www.openrent.co.uk/search/propertiesbyid?
ids=2052135](https://www.openrent.co.uk/search/propertiesbyid?ids=2052135)



What if we change the
number to another? (Pick
one that you know relates
to another property
if you can)

Remember this?

https://poo.gov.uk/wp-admin/admin-ajax.php?
action=update_tiles&
queryParams=%7B%22document_type%22%3A%22fii-report
%22%2C%22posts_per_page%22%3A50%2C%22paged%2
2%3A1%2C%22order%22%3A%22DESC%22%2C%22order
by%22%3A%22meta_value%22%2C%22meta_key%22%3A
%22document-date%22%7D
&_=1704708818980

Equals signs indicate parameters (**ingredients**)

The question mark indicates a **query**

Each ingredient is separated by an ampersand: &

Parts beginning with % are ways of expressing characters like “ in a URL:
%22 is “, %7B is {, %2C is comma

Payload: 'decode' params in URL (Chrome)

The screenshot shows the Chrome DevTools Network tab with the 'Payload' tab selected. Below it, the 'Query String Parameters' section is expanded, showing the following data:

- action:** update_tiles
- queryParams:** {"document_type": "fii-report", "posts_per_page": 50, "paged": 1, "order": "DESC", "orderby": "meta_value", "meta_key": "document-date"}
- _:** 1704872010964

Example from tinyurl.com/3mywvs3w
or ppo.gov.uk/document/fii-report/

Open the URL and change the parameters...

[https://ppo.gov.uk/wp-admin/admin-ajax.php?
action=update_tiles&
queryParams=%7B%22document_type%22%3A%22fii-report%22%2C%22posts_per_page%22%3A**100**%2C%22paged%22%3A1%2C%22order%22%3A%22DESC%22%2C%22orderby%22%3A%22meta_value%22%2C%22meta_key%22%3A%22document-date%22%7D
&_=1704708818980](https://ppo.gov.uk/wp-admin/admin-ajax.php?action=update_tiles&queryParams=%7B%22document_type%22%3A%22fii-report%22%2C%22posts_per_page%22%3A100%2C%22paged%22%3A1%2C%22order%22%3A%22DESC%22%2C%22orderby%22%3A%22meta_value%22%2C%22meta_key%22%3A%22document-date%22%7D&_=1704708818980)

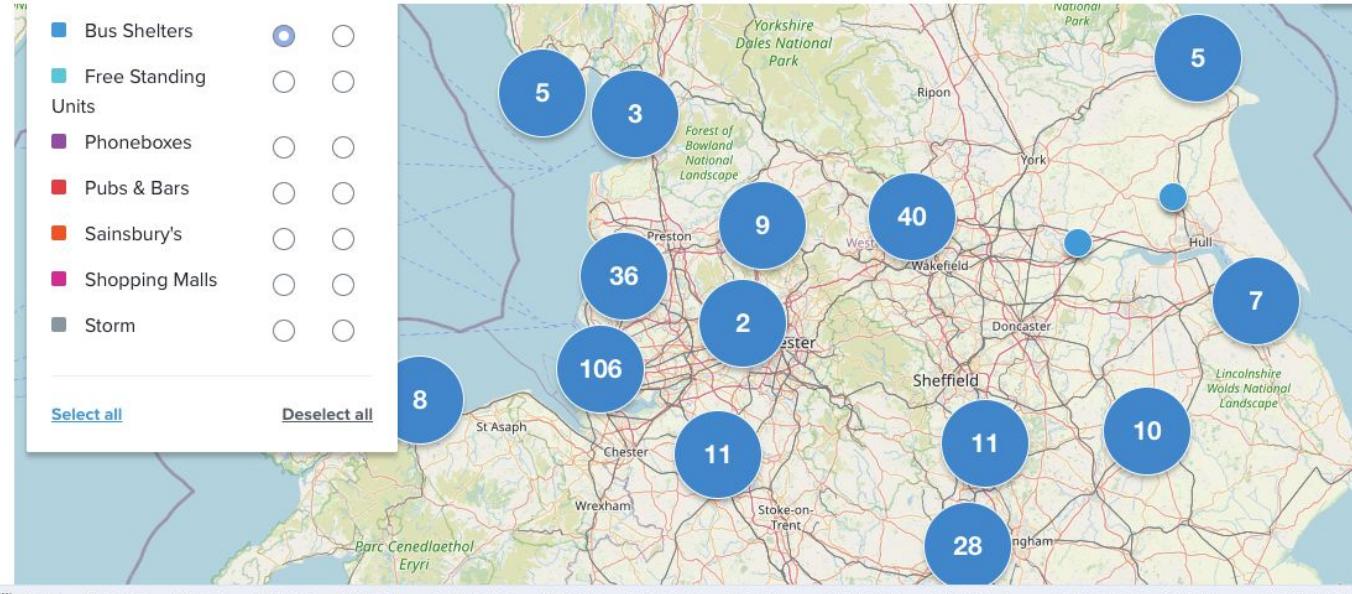
Check out Amazon's autosuggest API

Keep the Inspector's Network tab open while you type in the search box and watch the calls to the API

- URLs begin with **suggestions?limit=11**
- When opened in a new tab, full URL begins
https://completion.amazon.co.uk/api/2017/
- A 'call' is made to the API after each letter is typed: c, co, cof, etc.
- Break down Payload/URL: page-type=Gateway - are there different results on other pages? (E.g. Top Deals)

Other URLs that use undocumented APIs

- [Ticketmaster ticket listings](#)
- [Clear Channel campaign builder](#)
- [Thomas Cook holidays](#)
- [Sainsburys Store Locator](#)
- [Woolworths \(NZ\) store finder](#)
- [Vinted search results](#)
- [Clear Channel advertising sites map](#)
- [Fantasy Premier League](#)



Elements Console Sources **Network** Performance Memory Application Security Lighthouse Recorder ⚡ Performance insights ⚡ Web Scraper

Preserve log Disable cache No throttling

Filter Invert Hide data URLs Hide extension URLs All Fetch/XHR JS CSS Img Media Font Doc WS Wasm Manifest Other Blocked response

Name

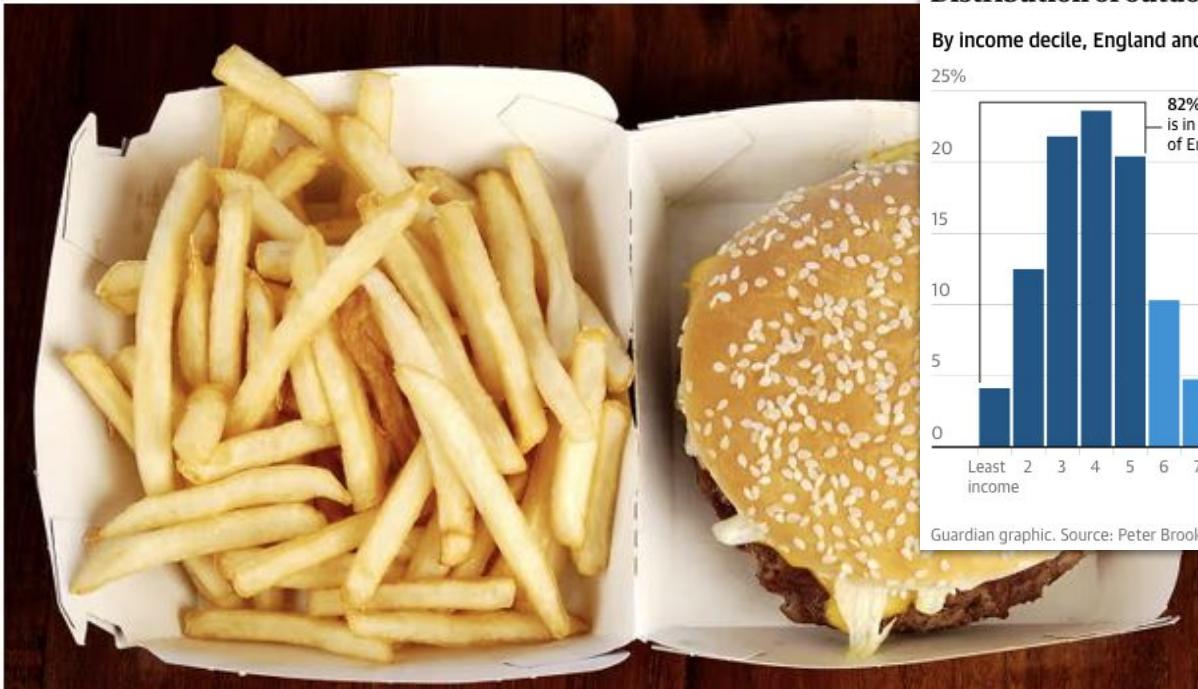
sites?duration=1&gender=all&age=all&types%5B0%5D%5...5B8%5D%5Bdigital%5D=1&types%...
 consentrecepts
 60514
 53626
 sites?duration=1&gender=all&age=all&types%5B0%5D%5...5B8%5D%5Bdigital%5D=1&types%...
 sites?duration=1&gender=all&age=all&types%5B0%5D%5...5B8%5D%5Bdigital%5D=0&types%...
 sites?duration=1&gender=all&age=all&types%5B0%5D%5...5B8%5D%5Bdigital%5D=0&types%...
 sites?duration=1&gender=all&age=all&types%5B0%5D%5...5B8%5D%5Bdigital%5D=0&types%...

Name	Type	Value
"sites?duration=1&gender=all&age=all&types%5B0%5D%5...5B8%5D%5Bdigital%5D=1&types%..."	JS	{"data": [
"consentrecepts"	JS	"59009,51.468059,-2.612833,2,1",
"60514"	JS	"59010,51.471202,-2.593003,2,1",
"53626"	JS	"59011,51.448445,-2.605983,2,1",
"sites?duration=1&gender=all&age=all&types%5B0%5D%5...5B8%5D%5Bdigital%5D=1&types%..."	JS	"59012,51.460993,-2.608397,2,1",
"sites?duration=1&gender=all&age=all&types%5B0%5D%5...5B8%5D%5Bdigital%5D=0&types%..."	JS	"59013,51.461478,-2.572846,2,1",
"sites?duration=1&gender=all&age=all&types%5B0%5D%5...5B8%5D%5Bdigital%5D=0&types%..."	JS	"59014,51.480087,-2.529574,2,1",
"sites?duration=1&gender=all&age=all&types%5B0%5D%5...5B8%5D%5Bdigital%5D=0&types%..."	JS	"59015,51.456635,-2.586965,2,1",
"sites?duration=1&gender=all&age=all&types%5B0%5D%5...5B8%5D%5Bdigital%5D=0&types%..."	JS	"59016,51.482713,-3.218662,2,1",
"sites?duration=1&gender=all&age=all&types%5B0%5D%5...5B8%5D%5Bdigital%5D=0&types%..."	JS	"59017,51.477050,-3.174390,2,1",
"sites?duration=1&gender=all&age=all&types%5B0%5D%5...5B8%5D%5Bdigital%5D=0&types%..."	JS	"59019,51.477709,-3.174090,2,1",

<https://www.clearchannel.co.uk/new-to-out-of-home/campaign-builder>

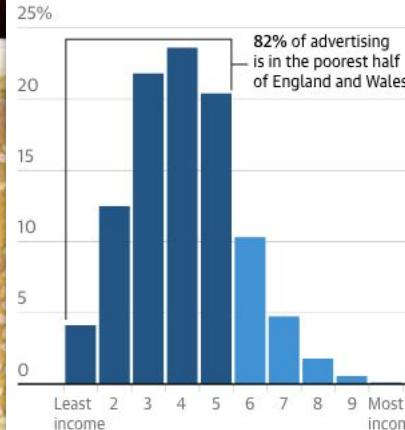
Four in five billboard ads in England and Wales in poorer areas

Exclusive: campaigners say outdoor ads pushing fast food, alcohol and more risk deepening health inequalities

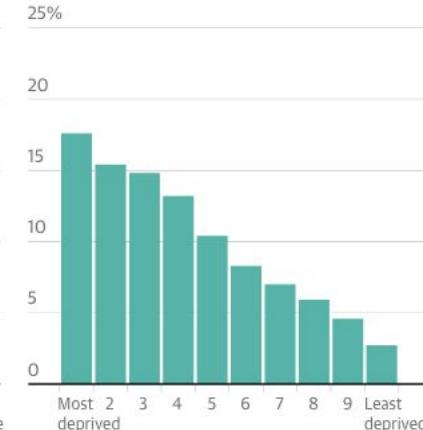


Distribution of outdoor advertising

By income decile, England and Wales



By deprivation decile, England



Guardian graphic. Source: Peter Brooks, Adfree Cities. Note: Net annual income after housing costs (ONS 2018)

[Clea Skopeliti, Guardian, 2024](#)
[AdFreeCities release; report](#)

Code and cURL

Using with Python: pandas and requests

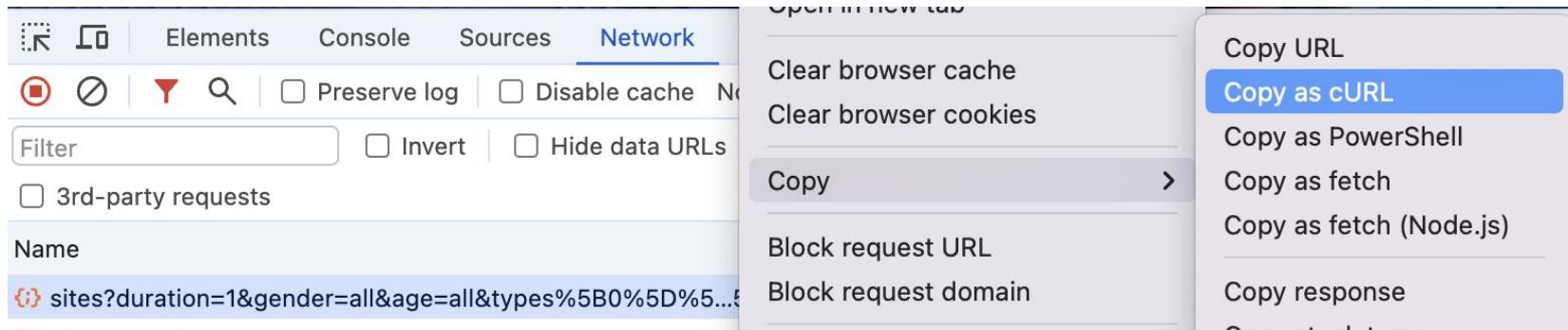
- If the JSON is simple then `pd.read_json()` should be able to take the URL and import the JSON into a pandas dataframe
- But multi-branching JSON will generate **ValueError: All arrays must be of the same length**
- Instead use **Requests** then parse the resulting object (drill down to the branch needed) and use `pd.DataFrame.from_dict()`. [This notebook](#) walks through the process

Using with R: jsonlite

- In R, you can import with **jsonlite::fromJSON()**
- To flatten branches:
`jsonlite::fromJSON(json_file, flatten = TRUE)`

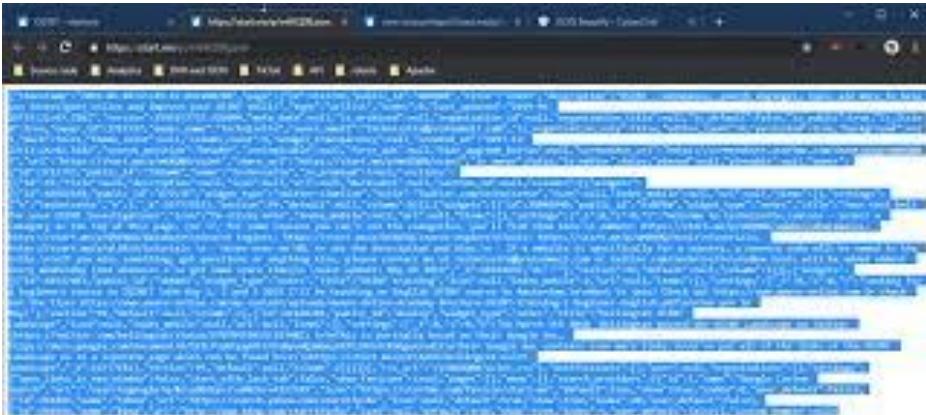
Using Curlconverter

- Right-click > Copy > Copy as cURL
- Paste into curlconverter.com to get Python/R
- Copy, paste into notebook and run



Find out more:

- Derek Willis: [Freeing the Plum Book](#)
- The MarkUp: [Journalists, You Should Be Looking for Undocumented APIs. Here's How to Start](#)
- Julia Evans: [How to use undocumented web APIs](#)
- [How can you find hidden or undocumented APIs for scraping?](#)
- [Scraping XHR](#)
- [The Lab #51: APIs with Bearer Token](#)



[10 Minute OSINT Tip: What Can We Learn From XHR, and JSON During an OSINT Investigation?](#)
[HakByte: How to use Postman to Reverse Engineer Private APIs](#)

If you can't find data, try All (POST methods)

The screenshot shows the Network tab of a browser developer tools interface. A red box highlights the 'All' tab in the top navigation bar. Another red box highlights the 'Request' tab in the right-hand sidebar. A third red box highlights the 'POST' method and the lock icon in the list of network requests. A large red arrow points from the 'Request' sidebar to a box containing the 'Form data' parameters.

tat...	Met...	Domain	File	Initiator	Type	Transferred	Si:
200	POST	www.wedino...	sample-results	jquery.min.j...	html	7.95 kB	49.
200	GET	www.wedino...	45586.jpg		img	jpeg	80.57 kB
200	GET	www.wedino...	45585.jpg		img	jpeg	84.37 kB
200	GET	www.wedino...	45583.jpg		img	jpeg	83.08 kB
200	GET	www.wedino...	45582.jpg		img	jpeg	84.29 kB
200	GET	www.wedino...	45581.jpg		img	jpeg	83.69 kB
200	GET	www.wedino...	45580.jpg		img	jpeg	81.87 kB
200	GET	www.wedino...	45579.jpg		img	jpeg	81.29 kB
200	GET	www.wedino...	45577.jpg		img	jpeg	86.10 kB
200	GET	www.wedino...	45575.jpg		img	jpeg	84.48 kB
200	GFT	www.wedino...	45571.ina		ima	ineq	83.52 kB

Form data

```
SD: "2024-05-05"  
ED: "2024-05-11"  
pageno: "2"  
RadioSearch: "5"  
srchReference: ""  
srchLocation: "1"  
srchKeyword: ""
```

<https://www.wedinos.org/sample-results> (and click on 'next')