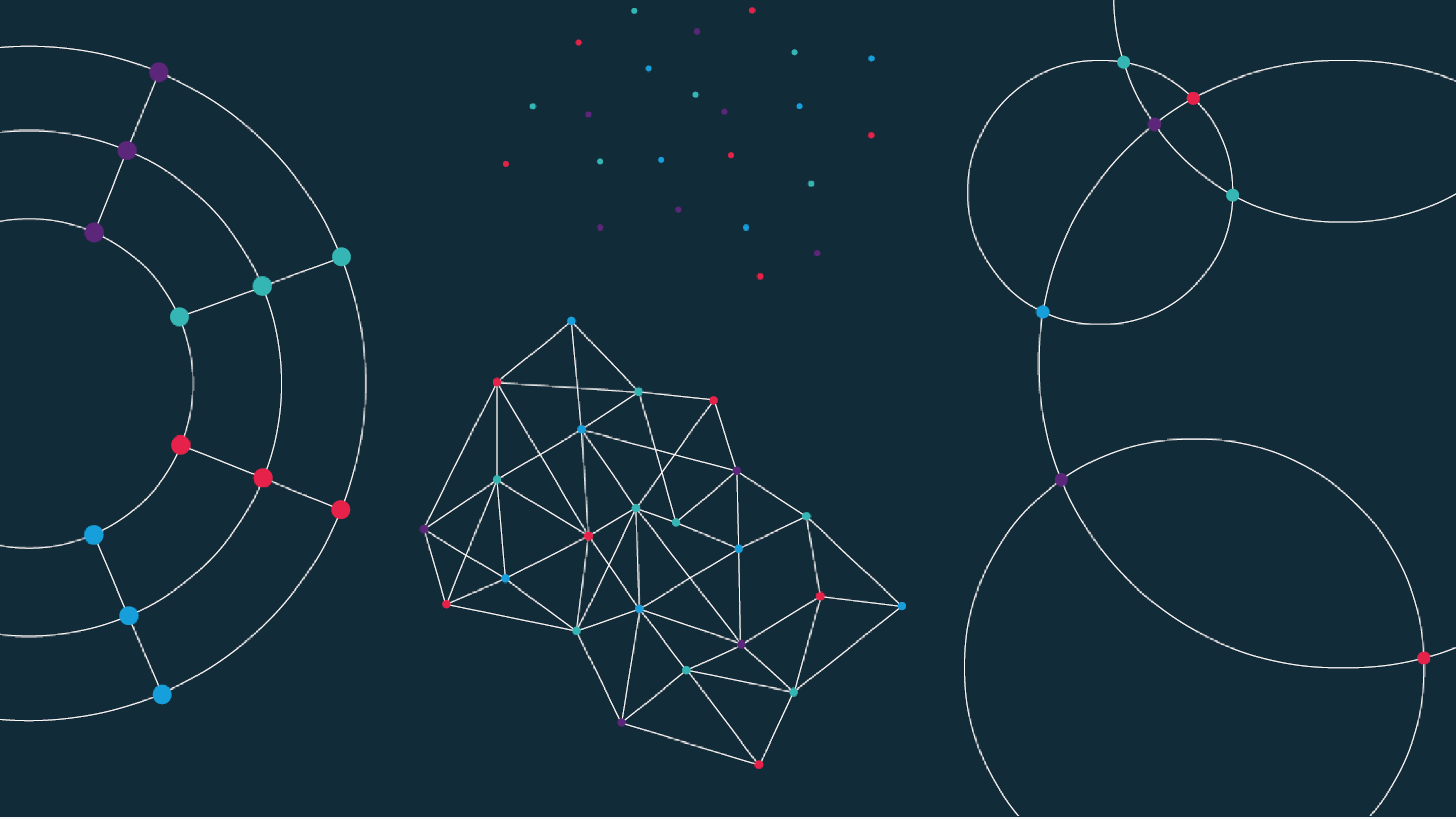
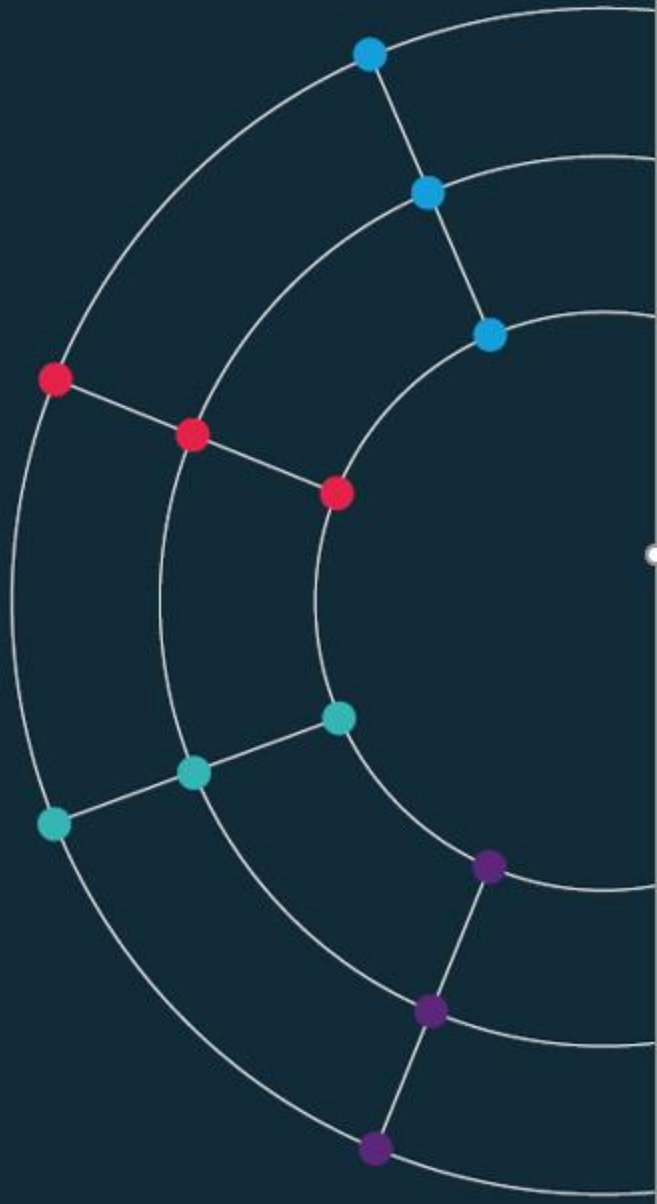


Session 5.

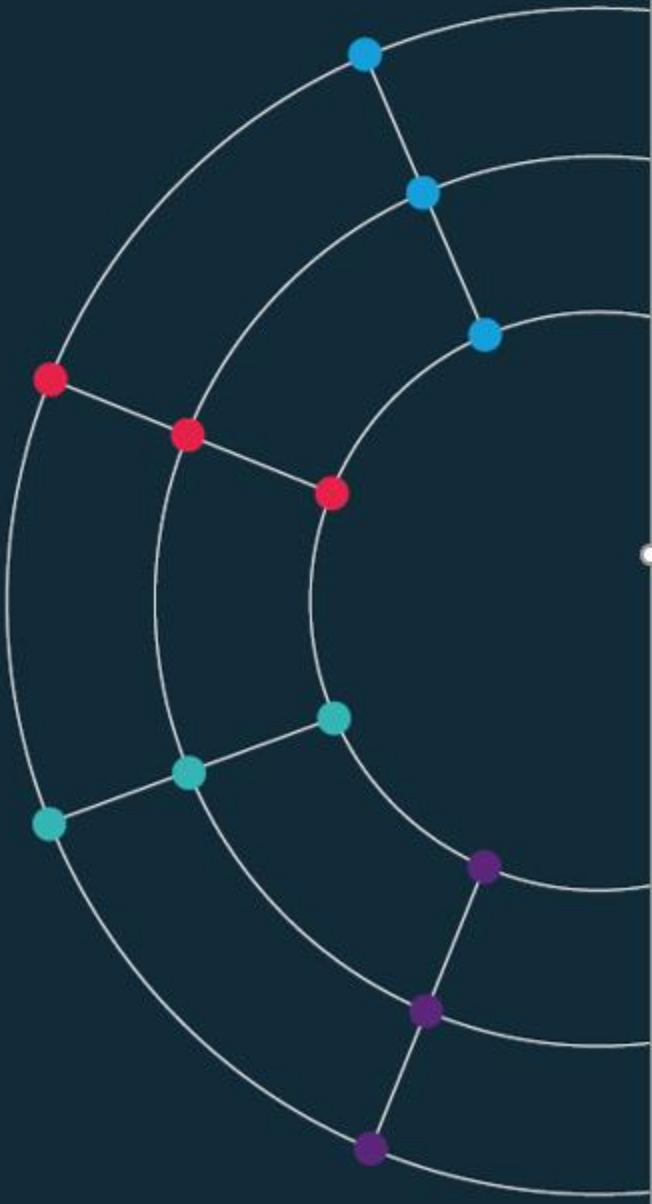
Data scraping



Session 5.

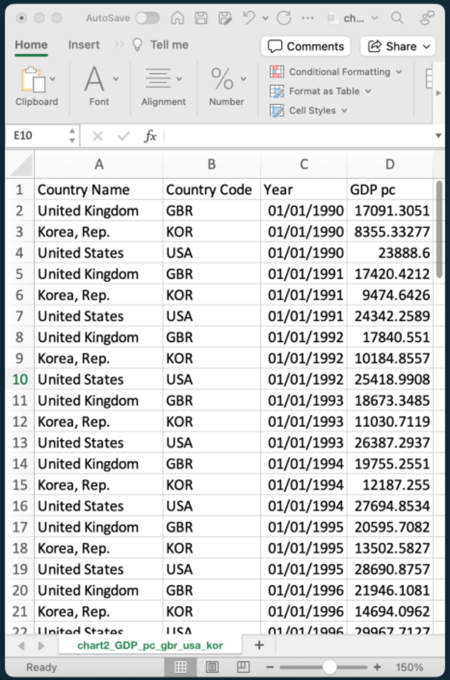
Data scraping

Scraping the HTML source



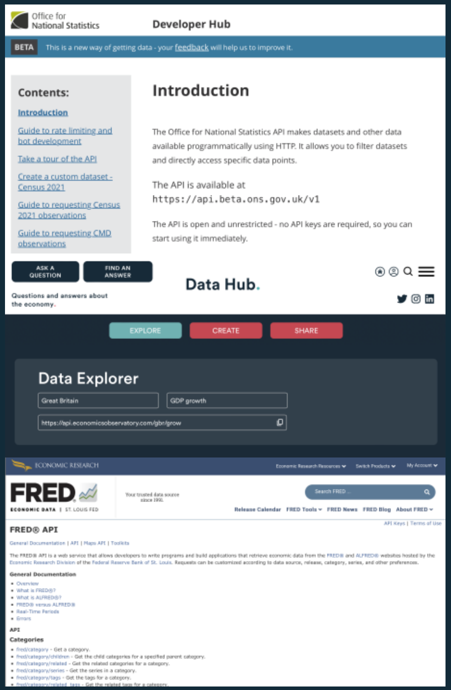
Data, so far...

Today we have used data from



	A	B	C	D
1	Country Name	Country Code	Year	GDP pc
2	United Kingdom	GBR	01/01/1990	17091.3051
3	Korea, Rep.	KOR	01/01/1990	8355.33277
4	United States	USA	01/01/1990	23888.6
5	United Kingdom	GBR	01/01/1991	17420.4212
6	Korea, Rep.	KOR	01/01/1991	9474.6426
7	United States	USA	01/01/1991	24342.2589
8	United Kingdom	GBR	01/01/1992	17840.551
9	Korea, Rep.	KOR	01/01/1992	10184.8557
10	United States	USA	01/01/1992	25418.9908
11	United Kingdom	GBR	01/01/1993	18673.3485
12	Korea, Rep.	KOR	01/01/1993	11030.7119
13	United States	USA	01/01/1993	26387.2937
14	United Kingdom	GBR	01/01/1994	19755.2551
15	Korea, Rep.	KOR	01/01/1994	12187.255
16	United States	USA	01/01/1994	27694.8534
17	United Kingdom	GBR	01/01/1995	20595.7082
18	Korea, Rep.	KOR	01/01/1995	13502.5827
19	United States	USA	01/01/1995	28690.8757
20	United Kingdom	GBR	01/01/1996	21946.1081
21	Korea, Rep.	KOR	01/01/1996	14694.0962
22	United States	USA	01/01/1996	29967.7127

Structured Files
(e.g. Excel, CSV, JSON)


Office for National Statistics
Developer Hub
BETA
Contents:
Introduction
Guide to rate limiting and bot development
Take a tour of the API
Create a custom dataset - Census 2021
Guide to requesting Census 2021 observations
Guide to requesting CMD observations
ASK A QUESTION
FIND AN ANSWER
Data Hub.
Questions and answers about the economy.
EXPLORE
CREATE
SHARE
Data Explorer
ECONOMIC RESEARCH
Economic Research Resources
Switch Products
My Account
FRED
ECONOMIC DATA | ST. LOUIS FED
Your trusted data source since 1991.
Search FRED ...
Release Calendar
FRED Tools
FRED News
FRED Blog
About FRED
API Keys | Terms of Use
General Documentation | API | Maps API | Toolkits

This is a new way of getting data - your feedback will help us to improve it.

Introduction

The Office for National Statistics API makes datasets and other data available programmatically using HTTP. It allows you to filter datasets and directly access specific data points.

The API is available at
`https://api.beta.ons.gov.uk/v1`

The API is open and unrestricted - no API keys are required, so you can start using it immediately.

Great Britain

GDP growth

https://api.economicsobservatory.com/gbr/grow

FRED® API

The FRED® API is a web service that allows developers to write programs and build applications that retrieve economic data from the FRED® and ALFRED® websites hosted by the Economic Research Division of the Federal Reserve Bank of St. Louis. Requests can be customized according to data source, release, category, series, and other preferences.

General Documentation

- Overview
- What is FRED®?
- What is ALFRED®?
- FRED® versus ALFRED®
- Real-Time Periods
- Errors

API

Categories

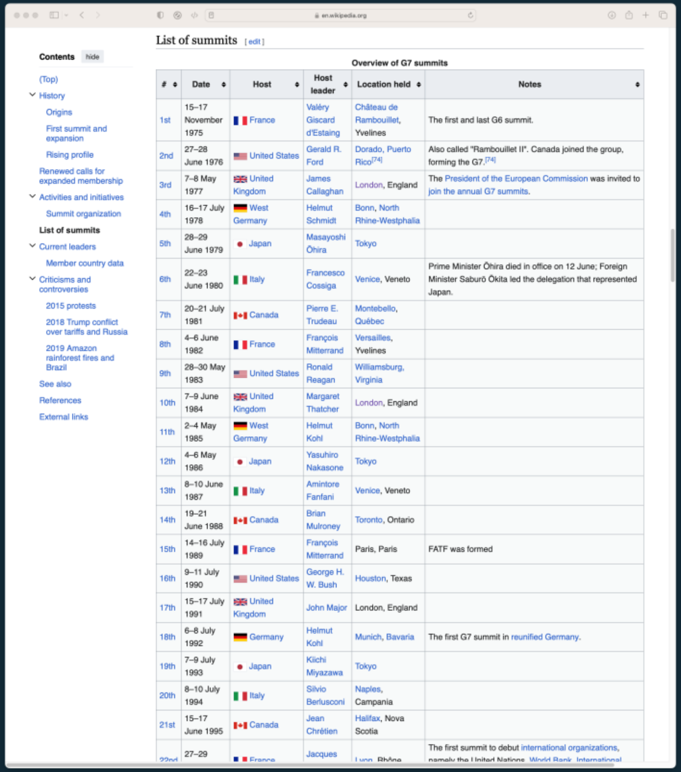
- fred/category - Get a category.
- fred/category/children - Get the child categories for a specified parent category.
- fred/category/related - Get the related categories for a category.
- fred/category/series - Get the series in a category.
- fred/category/tags - Get the tags for a category.
- fred/category/related_tags - Get the related tags for a category.

APIs
(e.g. ONS, ECO, FRED)

But what do we do when the data we want isn't available?

Data, so far...

What if we want



en.wikipedia.org

Contents hide

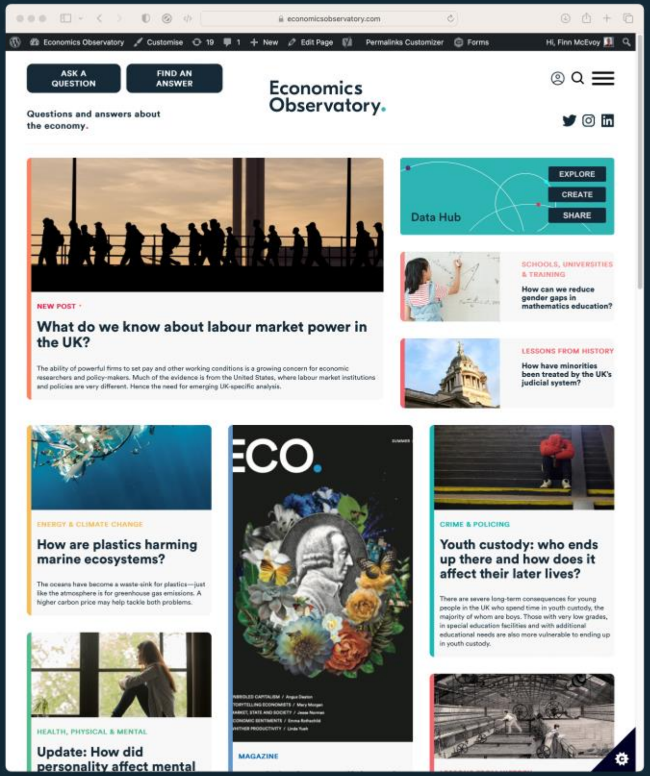
(Top)
- History
  - Origins
  - First summit and expansion
  - Rising profile
- Renewed calls for expanded membership
- Activities and initiatives
  - Summit organization
- List of summits
- Current leaders
  - Member country data
- Criticisms and controversies
  - 2015 protests
  - 2018 Trump conflict over tariffs and Russia
  - 2019 Amazon rainforest fires and Brazil
- See also
- References
- External links

List of summits [edit]

Overview of G7 summits

#	Date	Host	Host leader	Location held	Notes
1st	15–17 November 1975	France	Valéry Giscard d'Estaing	Château de Rambouillet, Yvelines	The first and last G6 summit.
2nd	27–28 June 1976	United States	Gerald R. Ford	Dorado, Puerto Rico[74]	Also called "Rambouillet II". Canada joined the group, forming the G7.[74]
3rd	7–8 May 1977	United Kingdom	James Callaghan	London, England	The President of the European Commission was invited to join the annual G7 summits.
4th	16–17 July 1978	West Germany	Helmut Schmidt	Bonn, North Rhine-Westphalia	
5th	28–29 June 1979	Japan	Masayoshi Ōhira	Tokyo	
6th	22–23 June 1980	Italy	Francesco Cossiga	Venice, Veneto	Prime Minister Ōhira died in office on 12 June; Foreign Minister Saburō Ōkita led the delegation that represented Japan.
7th	20–21 July 1981	Canada	Pierre E. Trudeau	Montebello, Québec	
8th	4–6 June 1982	France	François Mitterrand	Versailles, Yvelines	
9th	28–30 May 1983	United States	Ronald Reagan	Williamsburg, Virginia	
10th	7–9 June 1984	United Kingdom	Margaret Thatcher	London, England	
11th	2–4 May 1985	West Germany	Helmut Kohl	Bonn, North Rhine-Westphalia	
12th	4–6 May 1986	Japan	Yasuhiro Nakasone	Tokyo	
13th	8–10 June 1987	Italy	Amintore Fanfani	Venice, Veneto	
14th	19–21 June 1988	Canada	Brian Mulroney	Toronto, Ontario	
15th	14–16 July 1989	France	François Mitterrand	Paris, Paris	FATF was formed
16th	9–11 July 1990	United States	George H. W. Bush	Houston, Texas	
17th	15–17 July 1991	United Kingdom	John Major	London, England	
18th	6–8 July 1992	Germany	Helmut Kohl	Munich, Bavaria	The first G7 summit in reunified Germany.
19th	7–9 July 1993	Japan	Kiichi Miyazawa	Tokyo	
20th	8–10 July 1994	Italy	Silvio Berlusconi	Naples, Campania	
21st	15–17 June 1995	Canada	Jean Chrétien	Halifax, Nova Scotia	
22nd	27–29	France	Jacques	Lyon, Rhône	The first summit to debut international organizations, namely the United Nations, World Bank, International

Data from Wikipedia



economicsobservatory.com

Economics Observatory.

Questions and answers about the economy.

NEW POST

What do we know about labour market power in the UK?

The ability of powerful firms to set pay and other working conditions is a growing concern for economic researchers and policy-makers. Much of the evidence is from the United States, where labour market institutions and policies are very different. Hence the need for emerging UK-specific analysis.

SCHOOLS, UNIVERSITIES & TRAINING

How can we reduce gender gaps in mathematics education?

LESSONS FROM HISTORY

How have minorities been treated by the UK's judicial system?

ENERGY & CLIMATE CHANGE

How are plastics harming marine ecosystems?

The oceans have become a waste-sink for plastics—just like the atmosphere is for greenhouse gas emissions. A higher carbon price may help tackle both problems.

ECO.

MAGAZINE

CRIME & POLICING

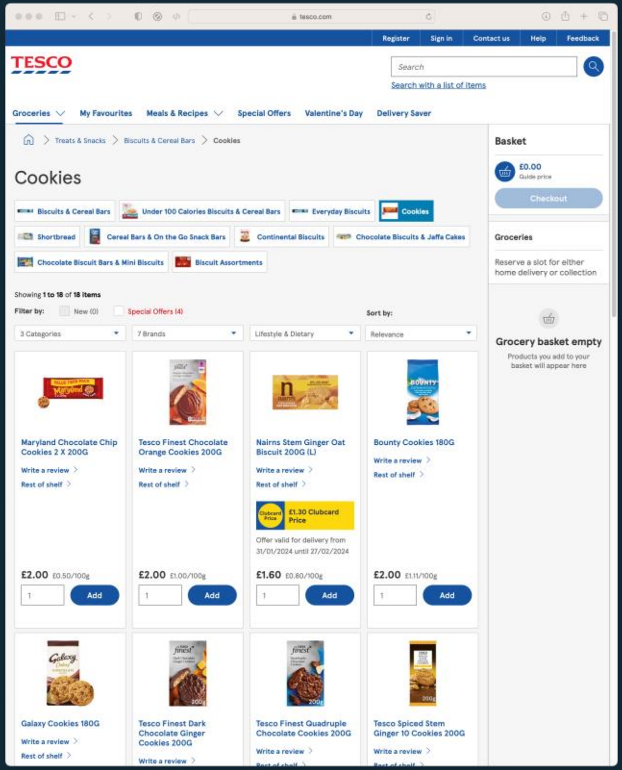
Youth custody: who ends up there and how does it affect their later lives?

There are severe long-term consequences for young people in the UK who spend time in youth custody, the majority of whom are boys. Those with very low grades, in special education facilities and with additional educational needs are also more vulnerable to ending up in youth custody.

HEALTH, PHYSICAL & MENTAL

Update: How did personality affect mental

News and Media



tesco.com

Cookies

Showing 1 to 18 of 18 items

Filter by: □ New (0) □ Special Offers (4)

3 Categories | 7 Brands | Lifestyle & Dietary | Sort by: Relevance

Basket

£0.00 Guide price

Groceries

Reserve a slot for either home delivery or collection

Grocery basket empty

Products you add to your basket will appear here

Product	Price	Unit price
Maryland Chocolate Chip Cookies 2 X 200G	£2.00	£0.50/100g
Tesco Finest Chocolate Orange Cookies 200G	£2.00	£1.00/100g
Nairns Stem Ginger Oat Biscuit 200G (L)	£1.60	£0.80/100g
Bounty Cookies 180G	£2.00	£1.11/100g

Nairns Stem Ginger Oat Biscuit 200G (L): Clubcard Price £1.30 Clubcard Price. Offer valid for delivery from 31/01/2024 until 27/02/2024

Galaxy Cookies 180G

Tesco Finest Dark Chocolate Ginger Cookies 200G

Tesco Finest Quadruple Chocolate Cookies 200G

Tesco Spiced Stem Ginger 10 Cookies 200G

Write a review

Rest of shelf

Prices from Supermarkets

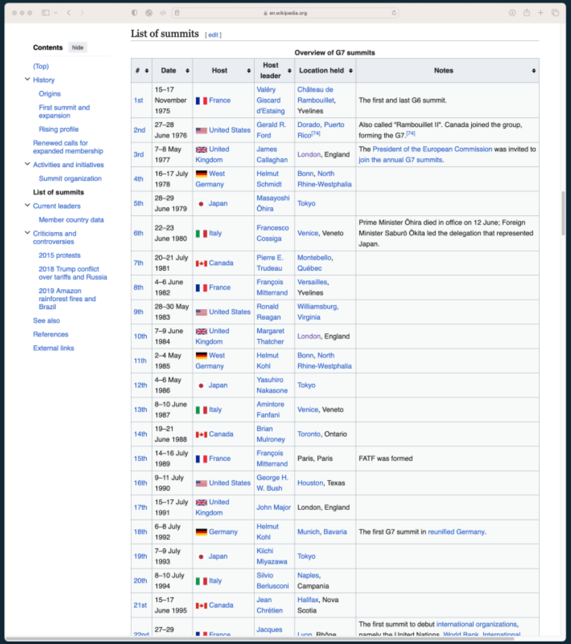
Scraping.

- The automated extraction of data from websites
- Scraping the HTML source. Easy. Automated.
- Scraping static HTML pages.
  - A bit more difficult. Can be automated.
  - CAPTCHA. Impersonating a human user. Zombie browser.
- Scraping HTML pages generated on-the-fly with JavaScript.
  - Hard. Only zombie browser works, and only in some cases.

hiQ labs vs. LinkedIn, 2019, US Court of Appeals for the Ninth Circuit, 17-16783
LinkedIn vs. hiQ labs, 2021, US Supreme Court, 19-1116

Your first scraper.

1. Extract data from Wikipedia



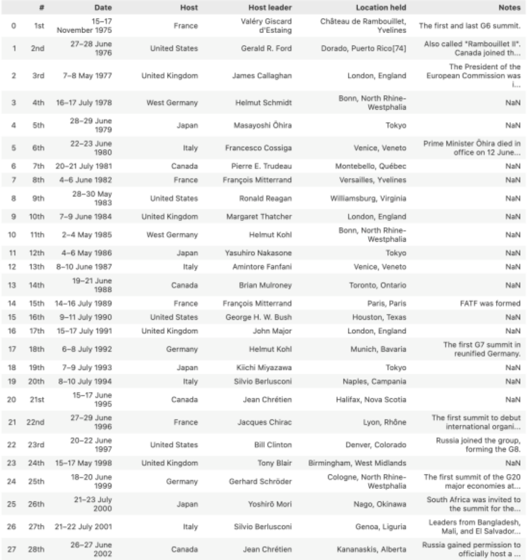
Contents hide

- (Top)
- History
  - Origins
  - First summit and expansion
  - Rising profile
- Renewed calls for expanded membership
- Activities and initiatives
  - Summit organization
- List of summits
- Current leaders
  - Member country data
- Criticisms and controversies
  - 2015 protests
  - 2018 Trump conflict over tariffs and Russia
  - 2019 Amazon rainforest fires and Brazil
- See also
- References
- External links

List of summits [edit]

Overview of G7 summits

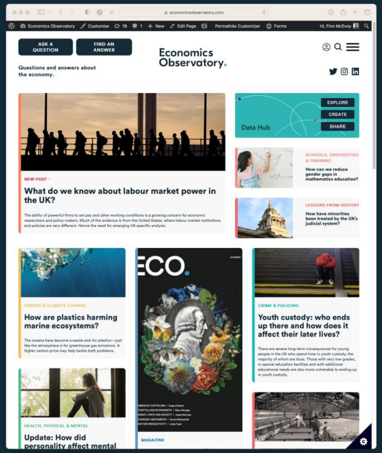
#	Date	Host	Host leader	Location held	Notes
1st	15–17 November 1975	France	Valéry Giscard d'Estaing	Château de Rambouillet, Yvelines	The first and last G6 summit.
2nd	27–28 June 1976	United States	Gerald R. Ford	Dorado, Puerto Rico[74]	Also called "Rambouillet II". Canada joined the group, forming the G7.[74]
3rd	7–8 May 1977	United Kingdom	James Callaghan	London, England	The President of the European Commission was invited to join the annual G7 summits.
4th	16–17 July 1978	West Germany	Helmut Schmidt	Bonn, North Rhine-Westphalia	
5th	28–29 June 1979	Japan	Masayoshi Ōhira	Tokyo	
6th	22–23 June 1980	Italy	Francesco Cossiga	Venice, Veneto	Prime Minister Ōhira died in office on 12 June; Foreign Minister Saburō Ōkita led the delegation that represented Japan.
7th	20–21 July 1981	Canada	Pierre E. Trudeau	Montebello, Québec	
8th	4–6 June 1982	France	François Mitterrand	Versailles, Yvelines	
9th	28–30 May 1983	United States	Ronald Reagan	Williamsburg, Virginia	
10th	7–9 June 1984	United Kingdom	Margaret Thatcher	London, England	
11th	2–4 May 1985	West Germany	Helmut Kohl	Bonn, North Rhine-Westphalia	
12th	4–6 May 1986	Japan	Yasuhiro Nakasone	Tokyo	
13th	8–10 June 1987	Italy	Amintore Fanfani	Venice, Veneto	
14th	19–21 June 1988	Canada	Brian Mulroney	Toronto, Ontario	
15th	14–16 July 1989	France	François Mitterrand	Paris, Paris	FATF was formed
16th	9–11 July 1990	United States	George H. W. Bush	Houston, Texas	
17th	15–17 July 1991	United Kingdom	John Major	London, England	
18th	6–8 July 1992	Germany	Helmut Kohl	Munich, Bavaria	The first G7 summit in reunified Germany.
19th	7–9 July 1993	Japan	Kiichi Miyazawa	Tokyo	
20th	8–10 July 1994	Italy	Silvio Berlusconi	Naples, Campania	
21st	15–17 June 1995	Canada	Jean Chrétien	Halifax, Nova Scotia	
22nd	27–29	France	Jacques	Lyon, Rhône	The first summit to debut international organizations, namely the United Nations, World Bank, International



	#	Date	Host	Host leader	Location held	Notes
0	1st	15–17 November 1975	France	Valéry Giscard d'Estaing	Château de Rambouillet, Yvelines	The first and last G6 summit.
1	2nd	27–28 June 1976	United States	Gerald R. Ford	Dorado, Puerto Rico[74]	Also called "Rambouillet II". Canada joined th...
2	3rd	7–8 May 1977	United Kingdom	James Callaghan	London, England	The President of the European Commission was i...
3	4th	16–17 July 1978	West Germany	Helmut Schmidt	Bonn, North Rhine-Westphalia	NaN
4	5th	28–29 June 1979	Japan	Masayoshi Ōhira	Tokyo	NaN
5	6th	22–23 June 1980	Italy	Francesco Cossiga	Venice, Veneto	Prime Minister Ōhira died in office on 12 June...
6	7th	20–21 July 1981	Canada	Pierre E. Trudeau	Montebello, Québec	NaN
7	8th	4–6 June 1982	France	François Mitterrand	Versailles, Yvelines	NaN
8	9th	28–30 May 1983	United States	Ronald Reagan	Williamsburg, Virginia	NaN
9	10th	7–9 June 1984	United Kingdom	Margaret Thatcher	London, England	NaN
10	11th	2–4 May 1985	West Germany	Helmut Kohl	Bonn, North Rhine-Westphalia	NaN
11	12th	4–6 May 1986	Japan	Yasuhiro Nakasone	Tokyo	NaN
12	13th	8–10 June 1987	Italy	Amintore Fanfani	Venice, Veneto	NaN
13	14th	19–21 June 1988	Canada	Brian Mulroney	Toronto, Ontario	NaN
14	15th	14–16 July 1989	France	François Mitterrand	Paris, Paris	FATF was formed
15	16th	9–11 July 1990	United States	George H. W. Bush	Houston, Texas	NaN
16	17th	15–17 July 1991	United Kingdom	John Major	London, England	NaN
17	18th	6–8 July 1992	Germany	Helmut Kohl	Munich, Bavaria	The first G7 summit in reunified Germany.
18	19th	7–9 July 1993	Japan	Kiichi Miyazawa	Tokyo	NaN
19	20th	8–10 July 1994	Italy	Silvio Berlusconi	Naples, Campania	NaN
20	21st	15–17 June 1995	Canada	Jean Chrétien	Halifax, Nova Scotia	NaN
21	22nd	27–29 June 1996	France	Jacques Chirac	Lyon, Rhône	The first summit to debut international organi...
22	23rd	20–22 June 1997	United States	Bill Clinton	Denver, Colorado	Russia joined the group, forming the G8.
23	24th	15–17 May 1998	United Kingdom	Tony Blair	Birmingham, West Midlands	NaN
24	25th	18–20 June 1999	Germany	Gerhard Schröder	Cologne, North Rhine-Westphalia	The first summit of the G20 major economies at...
25	26th	21–23 July 2000	Japan	Yoshirō Mori	Nago, Okinawa	South Africa was invited to the summit for the...
26	27th	21–22 July 2001	Italy	Silvio Berlusconi	Genoa, Liguria	Leaders from Bangladesh, Mali, and El Salvador...
27	28th	26–27 June 2002	Canada	Jean Chrétien	Kananaskis, Alberta	Russia gained permission to officially host a ...

Your first scraper.

2. Take a quick look at more complicated scraping



Questions and answers about the economy.

NEW POST

What do we know about labour market power in the UK?

The ability of powerful firms to set pay and other working conditions is a growing concern for economic researchers and policy-makers. Much of the evidence is from the United States, where labour market institutions and policies are very different. Hence the need for emerging UK-specific analysis.

SCHOOLS, UNIVERSITIES & TRAINING

How can we reduce gender gaps in mathematics education?

LESSONS FROM HISTORY

How have minorities been treated by the UK's judicial system?

ENERGY & CLIMATE CHANGE

How are plastics harming marine ecosystems?

The oceans have become a waste-sink for plastics—just like the atmosphere is for greenhouse gas emissions. A higher carbon price may help tackle both problems.

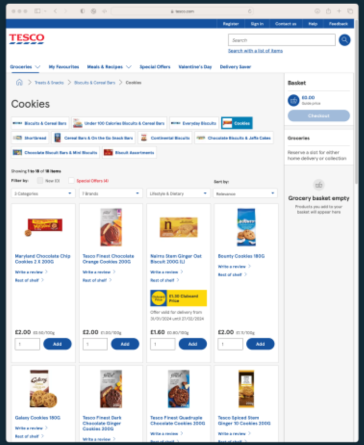
CRIME & POLICING

Youth custody: who ends up there and how does it affect their later lives?

There are severe long-term consequences for young people in the UK who spend time in youth custody, the majority of whom are boys. Those with very low grades, in special education facilities and with additional educational needs are also more vulnerable to ending up in youth custody.

HEALTH, PHYSICAL & MENTAL

Update: How did personality affect mental



Cookies

Showing 1 to 18 of 18 items

Maryland Chocolate Chip Cookies 2 X 200G
£2.00 £0.50/100g

Tesco Finest Chocolate Orange Cookies 200G
£2.00 £1.00/100g

Nairns Stem Ginger Oat Biscuit 200G (L)
£1.30 Clubcard Price
Offer valid for delivery from 31/01/2024 until 27/02/2024
£1.60 £0.80/100g

Bounty Cookies 180G
£2.00 £1.11/100g

Galaxy Cookies 180G

Tesco Finest Dark Chocolate Ginger Cookies 200G

Tesco Finest Quadruple Chocolate Cookies 200G

Tesco Spiced Stem Ginger 10 Cookies 200G

Reserve a slot for either home delivery or collection

Grocery basket empty

Products you add to your basket will appear here

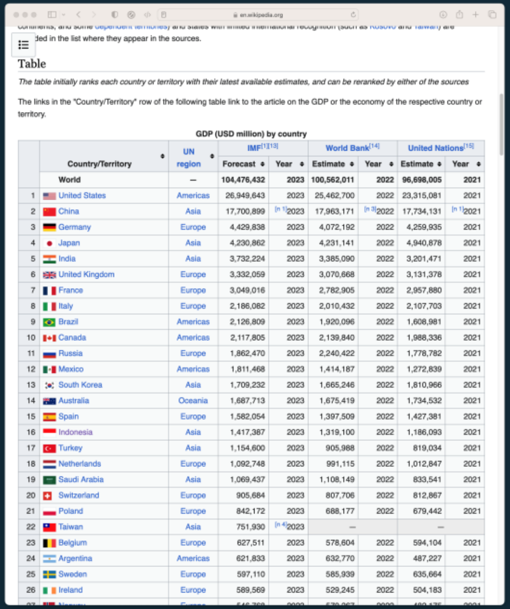
3. Share the tools to learn more

Overview.

1. Identify the data needed
2. Look Around – do you need to scrape?
3. Scrape the data
4. Cleaning and Visualising

Identify the data needed.

- You can scrape almost anything
- … but scraping is most useful for hard-to-find data



continents, and some dependent territories) and states with limited international recognition (such as Kosovo and Taiwan) are
ded in the list where they appear in the sources.

Table

The table initially ranks each country or territory with their latest available estimates, and can be reranked by either of the sources

The links in the "Country/Territory" row of the following table link to the article on the GDP or the economy of the respective country or territory.

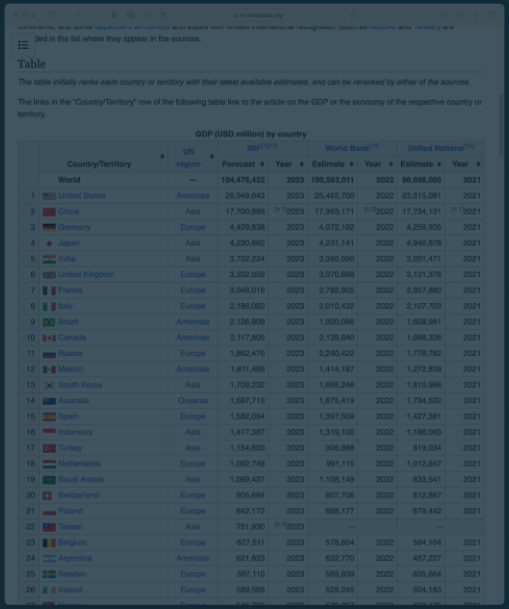
GDP (USD million) by country

	Country/Territory	UN region	IMF[1][13]		World Bank[14]		United Nations[15]	
			Forecast	Year	Estimate	Year	Estimate	Year
	World	—	104,476,432	2023	100,562,011	2022	96,698,005	2021
1	United States	Americas	26,949,643	2023	25,462,700	2022	23,315,081	2021
2	China	Asia	17,700,899	[n 1]2023	17,963,171	[n 3]2022	17,734,131	[n 1]2021
3	Germany	Europe	4,429,838	2023	4,072,192	2022	4,259,935	2021
4	Japan	Asia	4,230,862	2023	4,231,141	2022	4,940,878	2021
5	India	Asia	3,732,224	2023	3,385,090	2022	3,201,471	2021
6	United Kingdom	Europe	3,332,059	2023	3,070,668	2022	3,131,378	2021
7	France	Europe	3,049,016	2023	2,782,905	2022	2,957,880	2021
8	Italy	Europe	2,186,082	2023	2,010,432	2022	2,107,703	2021
9	Brazil	Americas	2,126,809	2023	1,920,096	2022	1,608,981	2021
10	Canada	Americas	2,117,805	2023	2,139,840	2022	1,988,336	2021
11	Russia	Europe	1,862,470	2023	2,240,422	2022	1,778,782	2021
12	Mexico	Americas	1,811,468	2023	1,414,187	2022	1,272,839	2021
13	South Korea	Asia	1,709,232	2023	1,665,246	2022	1,810,966	2021
14	Australia	Oceania	1,687,713	2023	1,675,419	2022	1,734,532	2021
15	Spain	Europe	1,582,054	2023	1,397,509	2022	1,427,381	2021
16	Indonesia	Asia	1,417,387	2023	1,319,100	2022	1,186,093	2021
17	Turkey	Asia	1,154,600	2023	905,988	2022	819,034	2021
18	Netherlands	Europe	1,092,748	2023	991,115	2022	1,012,847	2021
19	Saudi Arabia	Asia	1,069,437	2023	1,108,149	2022	833,541	2021
20	Switzerland	Europe	905,684	2023	807,706	2022	812,867	2021
21	Poland	Europe	842,172	2023	688,177	2022	679,442	2021
22	Taiwan	Asia	751,930	[n 4]2023	—		—	
23	Belgium	Europe	627,511	2023	578,604	2022	594,104	2021
24	Argentina	Americas	621,833	2023	632,770	2022	487,227	2021
25	Sweden	Europe	597,110	2023	585,939	2022	635,664	2021
26	Ireland	Europe	589,569	2023	529,245	2022	504,183	2021

You could scrape Wikipedia's list of countries by GDP…

Identify the data needed.

- You can scrape almost anything
- … but scraping is most useful for hard-to-find data



en.wikipedia.org

continents, and some dependent territories) and states with limited international recognition (such as Kosovo and Taiwan) are
ded in the list where they appear in the sources.

Table

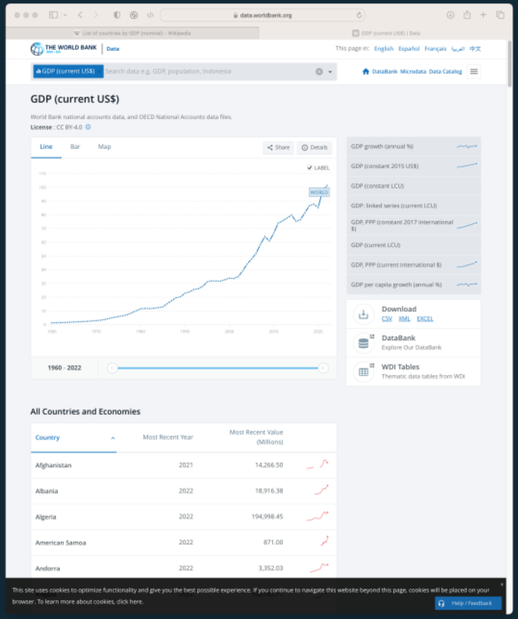
The table initially ranks each country or territory with their latest available estimates, and can be reranked by either of the sources

The links in the "Country/Territory" row of the following table link to the article on the GDP or the economy of the respective country or territory.

GDP (USD million) by country

	Country/Territory	UN region	IMF[1][13]		World Bank[14]		United Nations[15]	
			Forecast	Year	Estimate	Year	Estimate	Year
	World	—	104,476,432	2023	100,562,011	2022	96,698,005	2021
1	United States	Americas	26,949,643	2023	25,462,700	2022	23,315,081	2021
2	China	Asia	17,700,899	[n 1]2023	17,963,171	[n 3]2022	17,734,131	[n 1]2021
3	Germany	Europe	4,429,838	2023	4,072,192	2022	4,259,935	2021
4	Japan	Asia	4,230,862	2023	4,231,141	2022	4,940,878	2021
5	India	Asia	3,732,224	2023	3,385,090	2022	3,201,471	2021
6	United Kingdom	Europe	3,332,059	2023	3,070,668	2022	3,131,378	2021
7	France	Europe	3,049,016	2023	2,782,905	2022	2,957,880	2021
8	Italy	Europe	2,186,082	2023	2,010,432	2022	2,107,703	2021
9	Brazil	Americas	2,126,809	2023	1,920,096	2022	1,608,981	2021
10	Canada	Americas	2,117,805	2023	2,139,840	2022	1,988,336	2021
11	Russia	Europe	1,862,470	2023	2,240,422	2022	1,778,782	2021
12	Mexico	Americas	1,811,468	2023	1,414,187	2022	1,272,839	2021
13	South Korea	Asia	1,709,232	2023	1,665,246	2022	1,810,966	2021
14	Australia	Oceania	1,687,713	2023	1,675,419	2022	1,734,532	2021
15	Spain	Europe	1,582,054	2023	1,397,509	2022	1,427,381	2021
16	Indonesia	Asia	1,417,387	2023	1,319,100	2022	1,186,093	2021
17	Turkey	Asia	1,154,600	2023	905,988	2022	819,034	2021
18	Netherlands	Europe	1,092,748	2023	991,115	2022	1,012,847	2021
19	Saudi Arabia	Asia	1,069,437	2023	1,108,149	2022	833,541	2021
20	Switzerland	Europe	905,684	2023	807,706	2022	812,867	2021
21	Poland	Europe	842,172	2023	688,177	2022	679,442	2021
22	Taiwan	Asia	751,930	[n 4]2023	—		—	
23	Belgium	Europe	627,511	2023	578,604	2022	594,104	2021
24	Argentina	Americas	621,833	2023	632,770	2022	487,227	2021
25	Sweden	Europe	597,110	2023	585,939	2022	635,664	2021
26	Ireland	Europe	589,569	2023	529,245	2022	504,183	2021

You could scrape Wikipedia's list of countries by GDP…



data.worldbank.org

List of countries by GDP (nominal) - Wikipedia
GDP (current US$) | Data
THE WORLD BANK
Data
This page in: English Español Français العربية 中文
GDP (current US$)
Search data e.g. GDP, population, Indonesia
DataBank Microdata Data Catalog

GDP (current US$)

World Bank national accounts data, and OECD National Accounts data files.
License : CC BY-4.0

Line
Bar
Map
Share
Details
LABEL
WORLD
1960 - 2022

GDP growth (annual %)
GDP (constant 2015 US$)
GDP (constant LCU)
GDP: linked series (current LCU)
GDP, PPP (constant 2017 international $)
GDP (current LCU)
GDP, PPP (current international $)
GDP per capita growth (annual %)

Download
CSV XML EXCEL

DataBank
Explore Our DataBank

WDI Tables
Thematic data tables from WDI

All Countries and Economies

Country	Most Recent Year	Most Recent Value (Millions)
Afghanistan	2021	14,266.50
Albania	2022	18,916.38
Algeria	2022	194,998.45
American Samoa	2022	871.00
Andorra	2022	3,352.03

This site uses cookies to optimize functionality and give you the best possible experience. If you continue to navigate this website beyond this page, cookies will be placed on your browser. To learn more about cookies, click here.

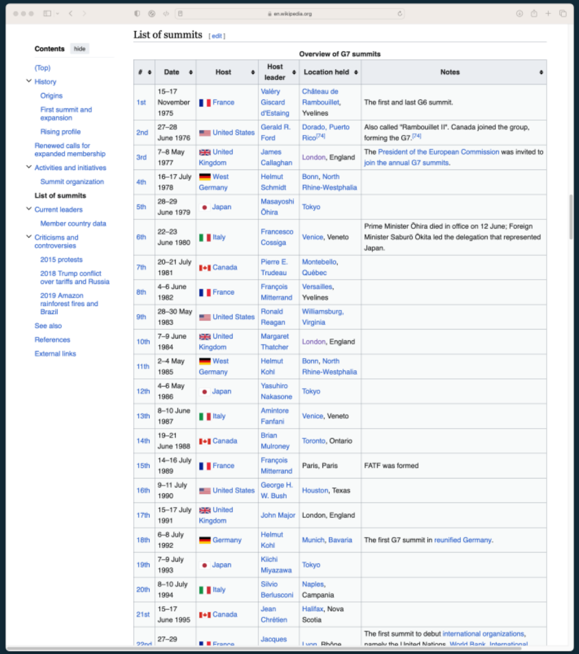
… but there's no point

If you can just download the data

Look around.

- Search the web for exactly the data you want

- Try to find a download first



Contents hide
(Top)
History
Origins
First summit and expansion
Rising profile
Renewed calls for expanded membership
Activities and initiatives
Summit organization
List of summits
Current leaders
Member country data
Criticisms and controversies
2015 protests
2018 Trump conflict over tariffs and Russia
2019 Amazon rainforest fires and Brazil
See also
References
External links

List of summits [edit]

Overview of G7 summits

#	Date	Host	Host leader	Location held	Notes
1st	15–17 November 1975	France	Valéry Giscard d'Estaing	Château de Rambouillet, Yvelines	The first and last G6 summit.
2nd	27–28 June 1976	United States	Gerald R. Ford	Dorado, Puerto Rico[74]	Also called "Rambouillet II". Canada joined the group, forming the G7.[74]
3rd	7–8 May 1977	United Kingdom	James Callaghan	London, England	The President of the European Commission was invited to join the annual G7 summits.
4th	16–17 July 1978	West Germany	Helmut Schmidt	Bonn, North Rhine-Westphalia	
5th	28–29 June 1979	Japan	Masayoshi Ōhira	Tokyo	
6th	22–23 June 1980	Italy	Francesco Cossiga	Venice, Veneto	Prime Minister Ōhira died in office on 12 June; Foreign Minister Saburō Ōkita led the delegation that represented Japan.
7th	20–21 July 1981	Canada	Pierre E. Trudeau	Montebello, Québec	
8th	4–6 June 1982	France	François Mitterrand	Versailles, Yvelines	
9th	28–30 May 1983	United States	Ronald Reagan	Williamsburg, Virginia	
10th	7–9 June 1984	United Kingdom	Margaret Thatcher	London, England	
11th	2–4 May 1985	West Germany	Helmut Kohl	Bonn, North Rhine-Westphalia	
12th	4–6 May 1986	Japan	Yasuhiro Nakasone	Tokyo	
13th	8–10 June 1987	Italy	Amintore Fanfani	Venice, Veneto	
14th	19–21 June 1988	Canada	Brian Mulroney	Toronto, Ontario	
15th	14–16 July 1989	France	François Mitterrand	Paris, Paris	FATF was formed
16th	9–11 July 1990	United States	George H. W. Bush	Houston, Texas	
17th	15–17 July 1991	United Kingdom	John Major	London, England	
18th	6–8 July 1992	Germany	Helmut Kohl	Munich, Bavaria	The first G7 summit in reunified Germany.
19th	7–9 July 1993	Japan	Kiichi Miyazawa	Tokyo	
20th	8–10 July 1994	Italy	Silvio Berlusconi	Naples, Campania	
21st	15–17 June 1995	Canada	Jean Chrétien	Halifax, Nova Scotia	
22nd	27–29	France	Jacques	Lyon, Rhône	The first summit to debut international organizations, namely the United Nations, World Bank, International

Our Example

Table of G7 Meetings

✓ Excel/CSV Unavailable

✓ Table Available

Scrape the data.

- Scraping data from tables on webpages is easy with Python
- We can use Pandas, which we have already seen today

```
pd.read_html(url)
```

(Loads every table from a webpage)

Scrape the data.

- To read all the tables, we point url to our example page

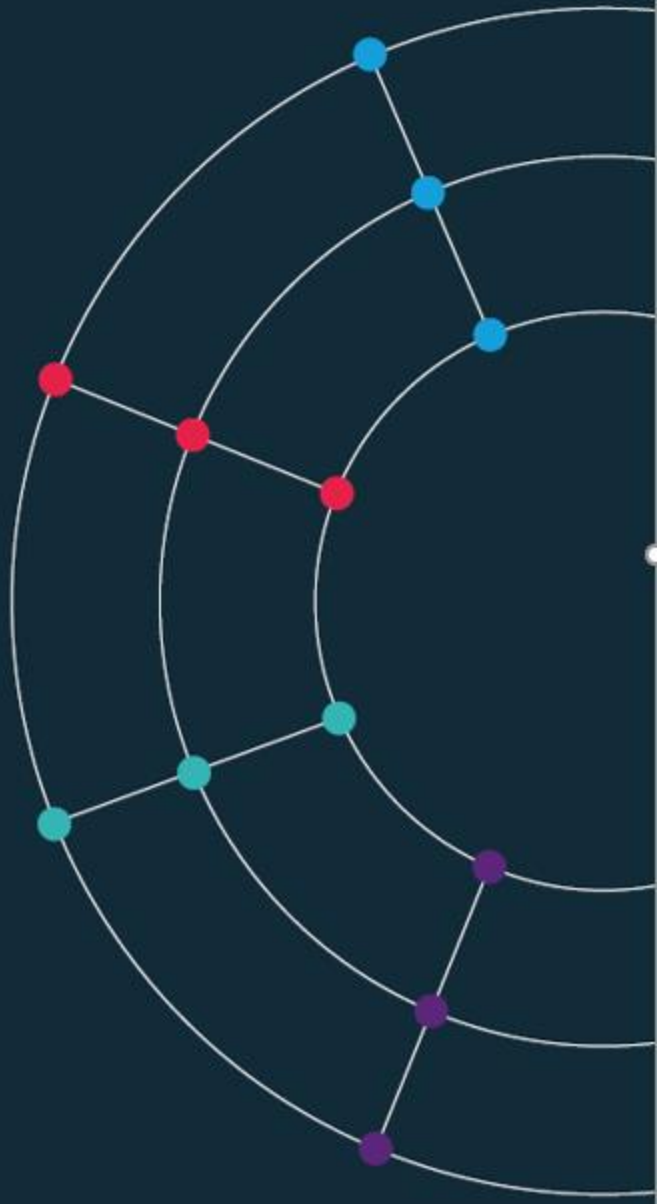
```
url = "https://en.wikipedia.org/wiki/G7"
tables_from_webpage = pd.read_html(url)
```

Which makes a list of every table.

Session 5.

Data scraping

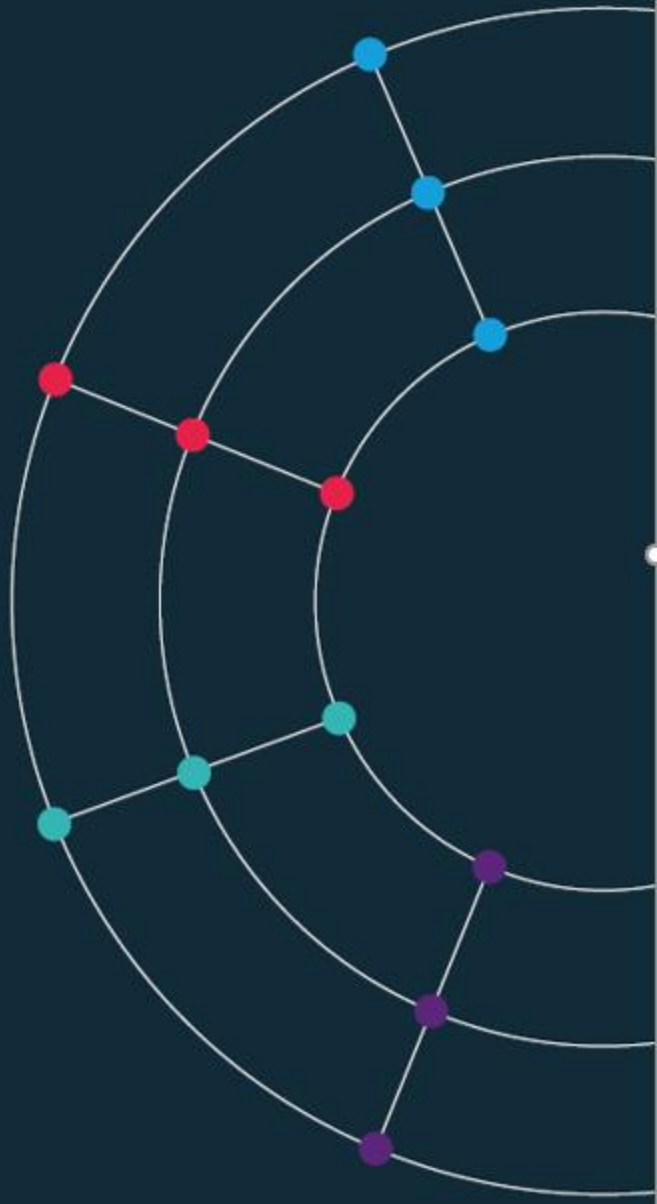
Scraping the HTML source



Session 5.

Data scraping

https://github.com/EconomicsObservatory/courses/blob/main/README.md



Code-along.

Your first scraper

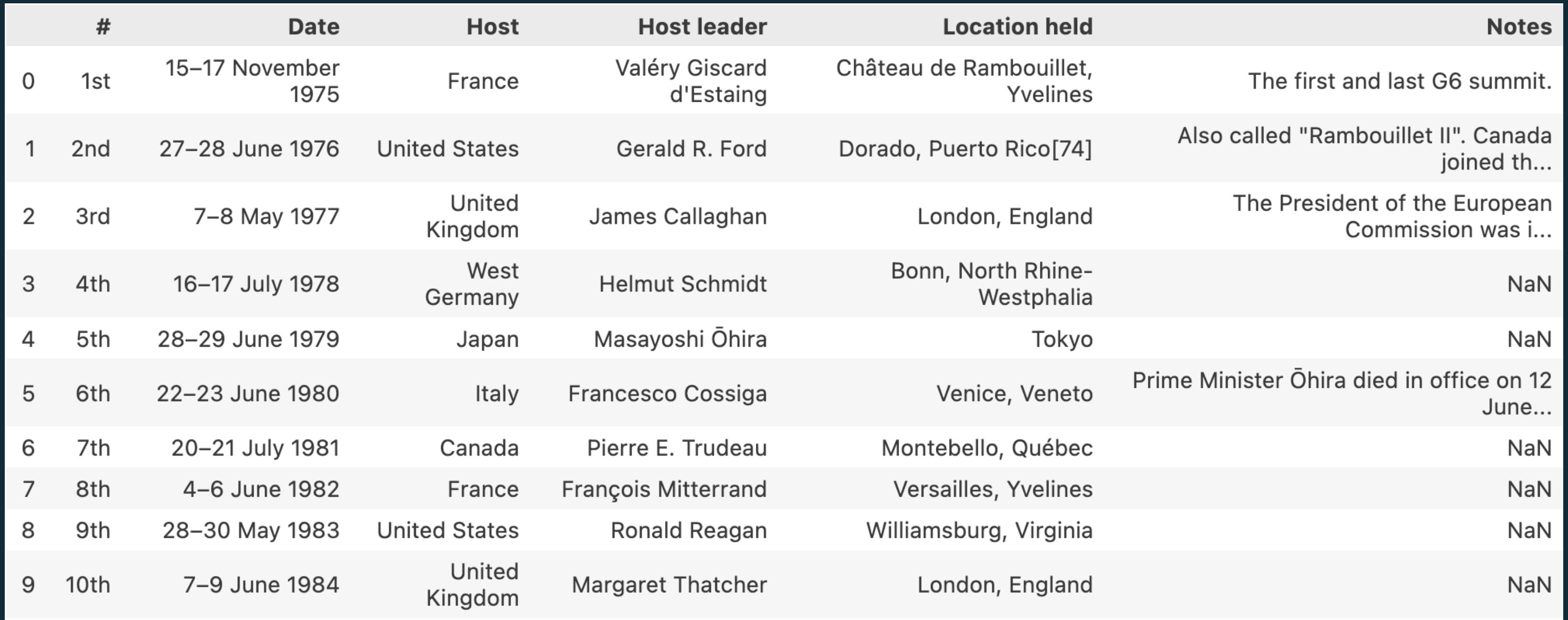
In this fifth practical session, we will use Google Colab to use Python to scape data from Wikipedia (and another example if there is time). We will also embed a chart displaying the scraped data into your website, using VS Code and GitHub.

- We will run you through the following guided notebook: “Session_5_Scraping_basic.ipynb” (open in Google Colab)
- For a further advanced examples, go to: “Session_5_Scraping_advanced.ipynb” (open in Google Colab)

Scrape the data.

The list of G7 meetings, our target, is the 3rd table on the webpage:

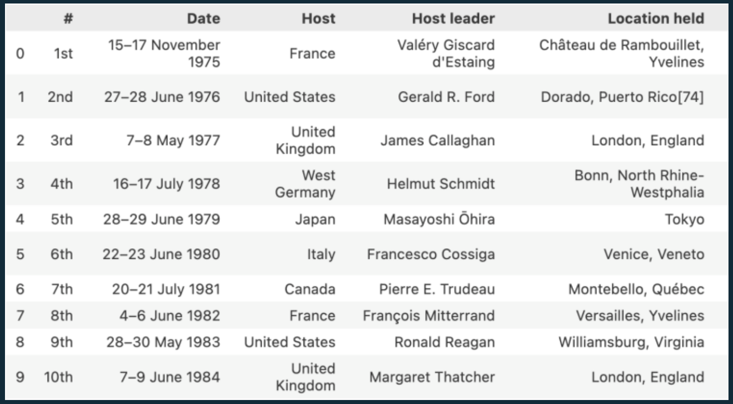
```
tables_from_webpage[2]
```



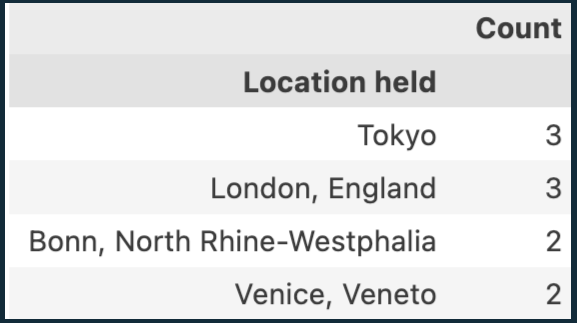
	#	Date	Host	Host leader	Location held	Notes
0	1st	15–17 November 1975	France	Valéry Giscard d'Estaing	Château de Rambouillet, Yvelines	The first and last G6 summit.
1	2nd	27–28 June 1976	United States	Gerald R. Ford	Dorado, Puerto Rico[74]	Also called "Rambouillet II". Canada joined th...
2	3rd	7–8 May 1977	United Kingdom	James Callaghan	London, England	The President of the European Commission was i...
3	4th	16–17 July 1978	West Germany	Helmut Schmidt	Bonn, North Rhine-Westphalia	NaN
4	5th	28–29 June 1979	Japan	Masayoshi Ōhira	Tokyo	NaN
5	6th	22–23 June 1980	Italy	Francesco Cossiga	Venice, Veneto	Prime Minister Ōhira died in office on 12 June...
6	7th	20–21 July 1981	Canada	Pierre E. Trudeau	Montebello, Québec	NaN
7	8th	4–6 June 1982	France	François Mitterrand	Versailles, Yvelines	NaN
8	9th	28–30 May 1983	United States	Ronald Reagan	Williamsburg, Virginia	NaN
9	10th	7–9 June 1984	United Kingdom	Margaret Thatcher	London, England	NaN

Cleaning and visualising.

- We have a messy table of data
- Let’s clean it up to answer the following question:
  - ‘What’s the most popular location for G7 meetings?’



	#	Date	Host	Host leader	Location held
0	1st	15–17 November 1975	France	Valéry Giscard d'Estaing	Château de Rambouillet, Yvelines
1	2nd	27–28 June 1976	United States	Gerald R. Ford	Dorado, Puerto Rico[74]
2	3rd	7–8 May 1977	United Kingdom	James Callaghan	London, England
3	4th	16–17 July 1978	West Germany	Helmut Schmidt	Bonn, North Rhine-Westphalia
4	5th	28–29 June 1979	Japan	Masayoshi Ōhira	Tokyo
5	6th	22–23 June 1980	Italy	Francesco Cossiga	Venice, Veneto
6	7th	20–21 July 1981	Canada	Pierre E. Trudeau	Montebello, Québec
7	8th	4–6 June 1982	France	François Mitterrand	Versailles, Yvelines
8	9th	28–30 May 1983	United States	Ronald Reagan	Williamsburg, Virginia
9	10th	7–9 June 1984	United Kingdom	Margaret Thatcher	London, England

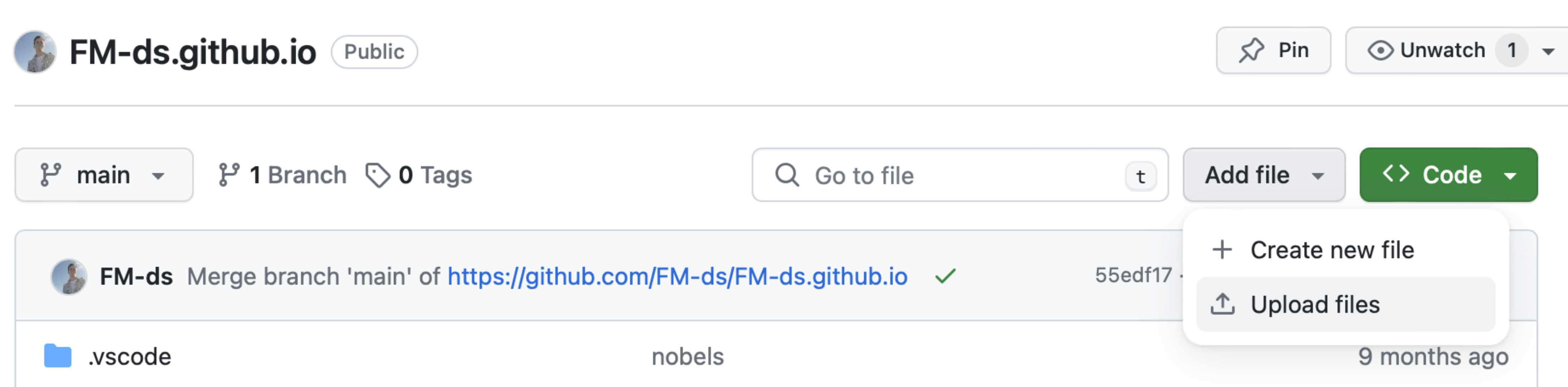


	Count
Location held	
Tokyo	3
London, England	3
Bonn, North Rhine-Westphalia	2
Venice, Veneto	2

(full code in Notebook)

Cleaning and visualising.

After saving our table and uploading to GitHub, we can use it in Vega-Lite

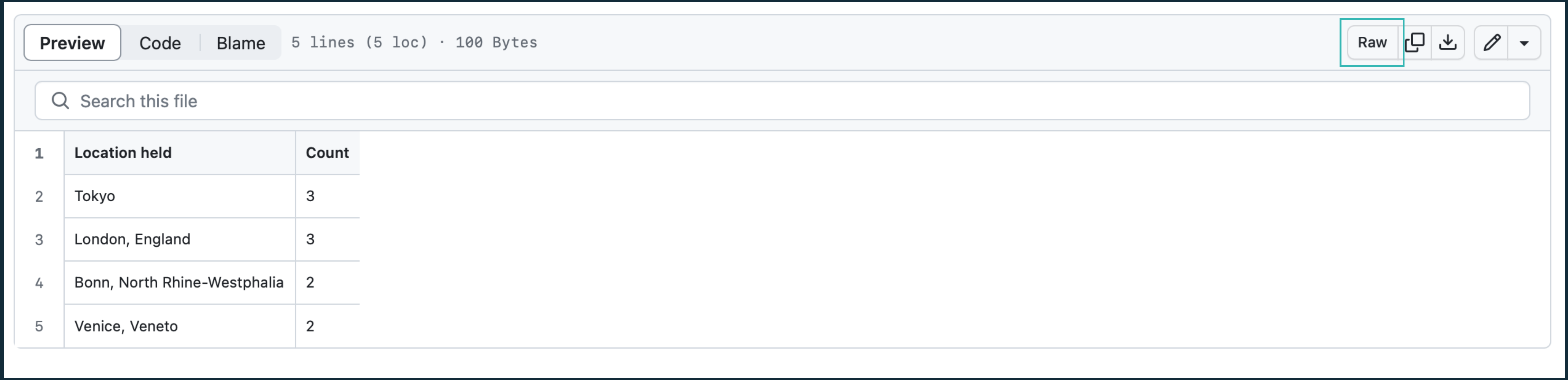


FM-ds Merge branch 'main' of https://github.com/FM-ds/FM-ds.github.io ✓ 55edf17

.vscode nobels 9 months ago

Cleaning and visualising.

After saving our table and uploading to GitHub, we can use it in Vega-Lite

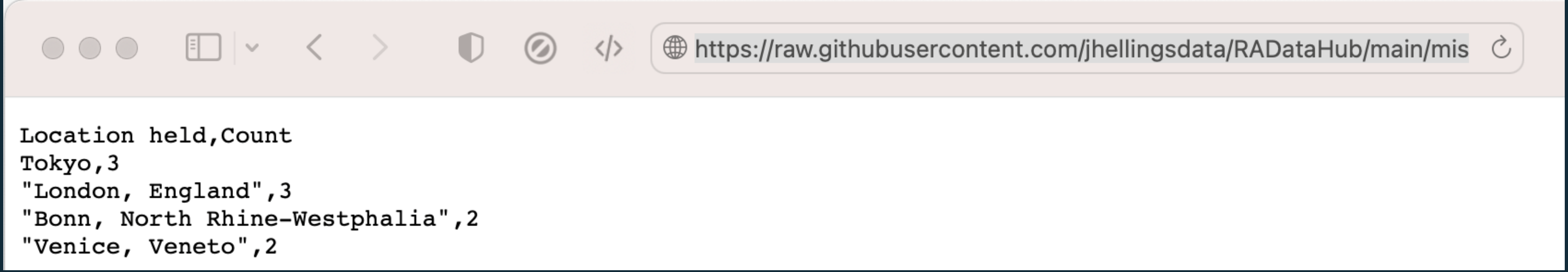


5 lines (5 loc) · 100 Bytes

Location held	Count
Tokyo	3
London, England	3
Bonn, North Rhine-Westphalia	2
Venice, Veneto	2

Cleaning and visualising.

After saving our table and uploading to GitHub, we can use it in Vega-Lite

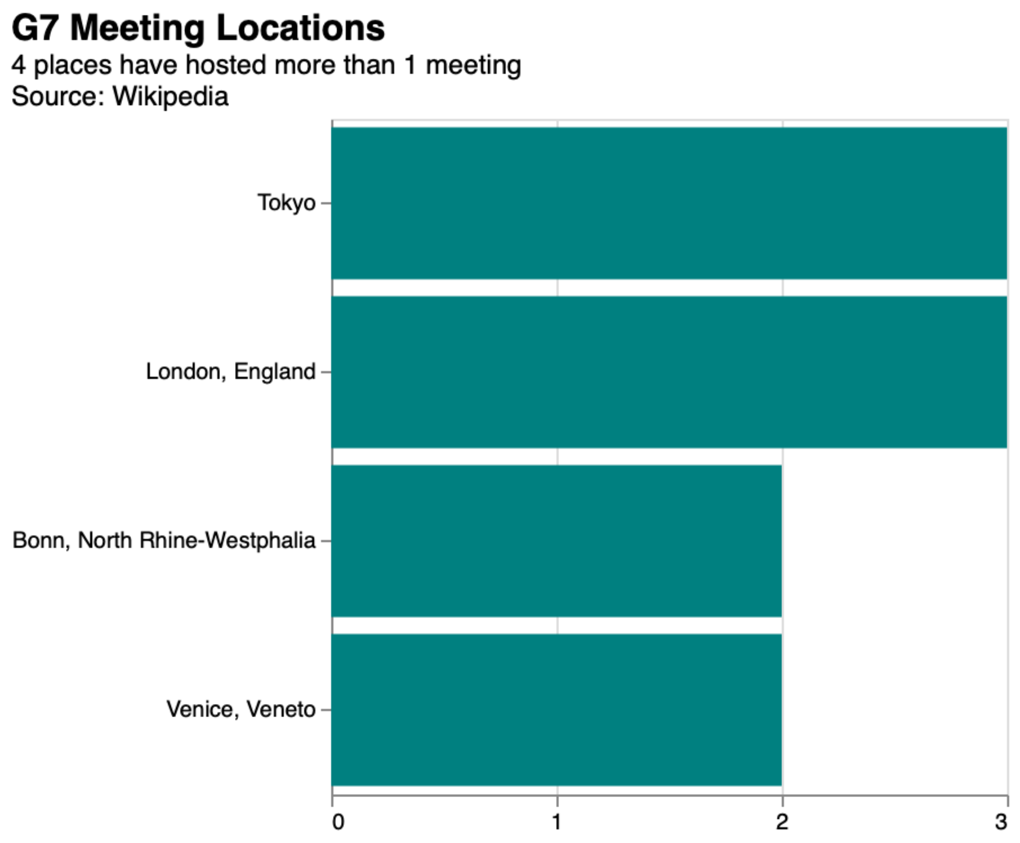


https://raw.githubusercontent.com/jhellingsdata/RADataHub/main/mis

```
Location held,Count
Tokyo,3
"London, England",3
"Bonn, North Rhine-Westphalia",2
"Venice, Veneto",2
```

Cleaning and visualising.

Linking to our data, we can use it in a chart:

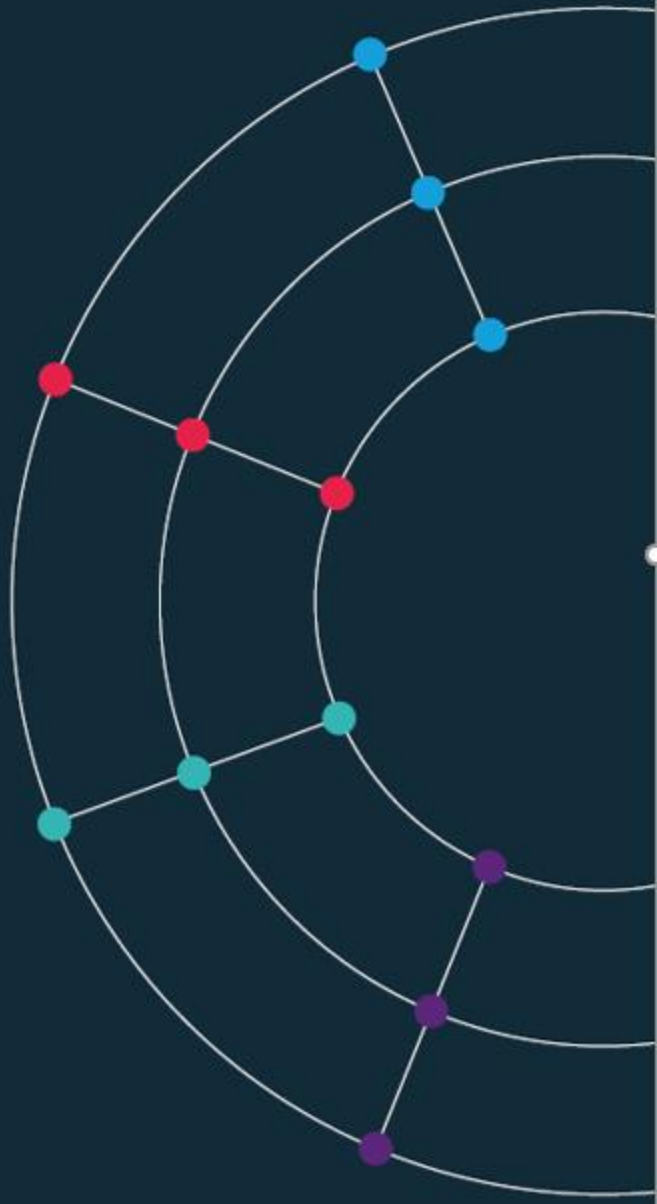


(see “chart_g7_meeting_hosts.json”)

Session 5.

Data scraping

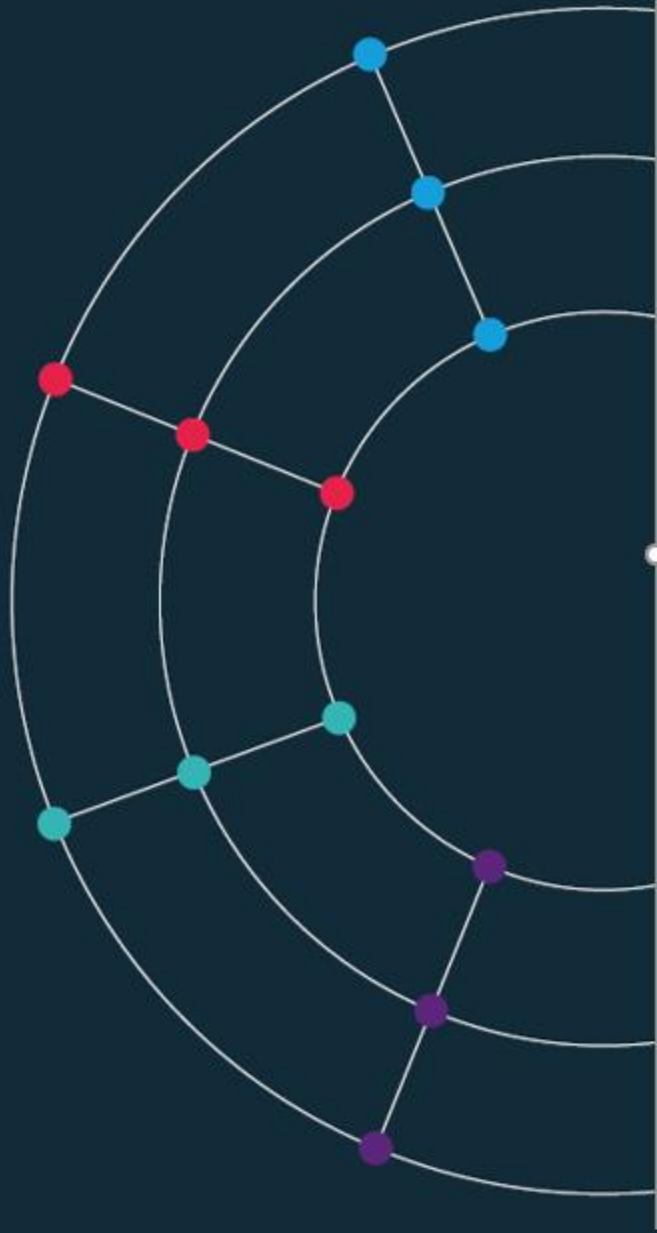
Scraping the HTML source (advanced)



Session 5.

Data scraping

https://github.com/EconomicsObservatory/courses/blob/main/README.md

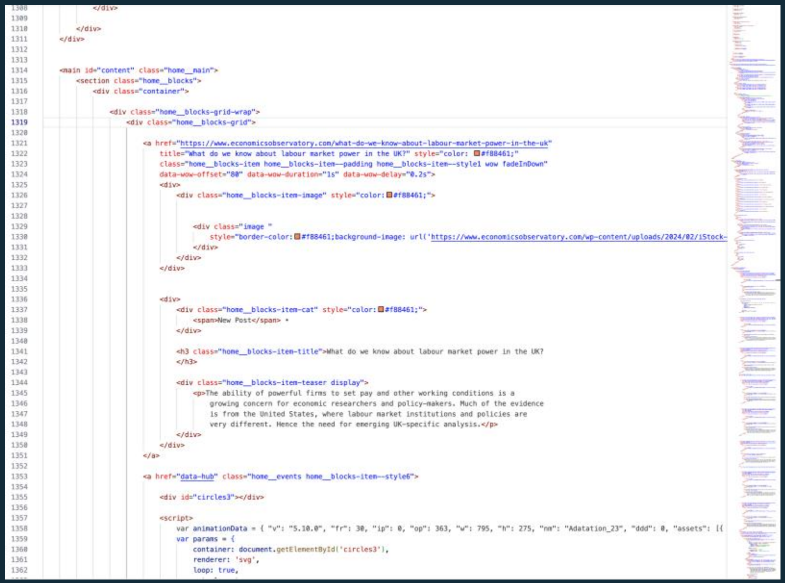


Scraping HTML Source.

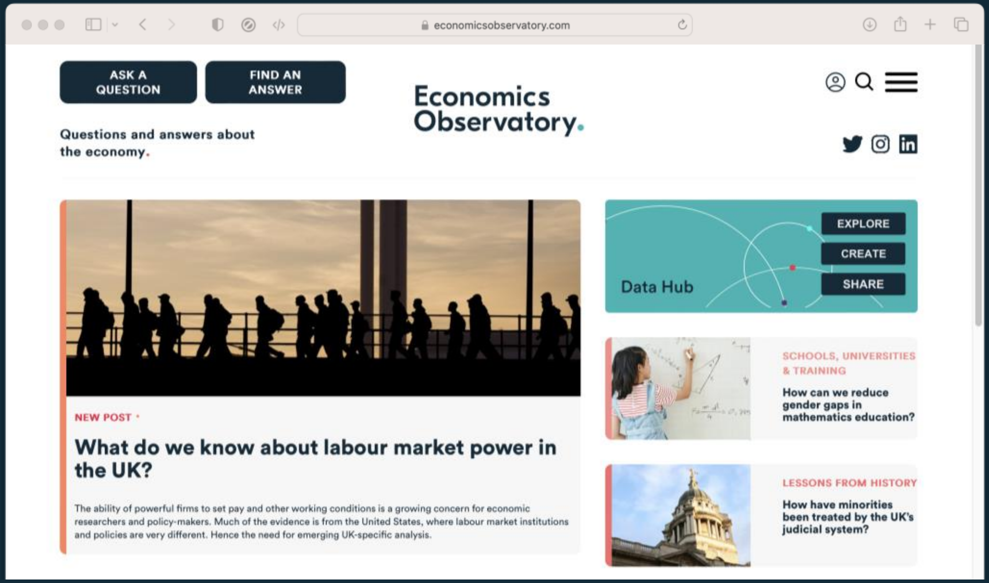
- Scraping tables is easy but sometimes we want data that isn't nicely formatted
- Instead, we can extract data by searching the HTML
- Everything is defined in the HTML, we just have to find it

Scraping HTML Source.

For example, we can scrape ECO headlines and tag-lines by ‘parsing’ the HTML



```
        </div>

    </div>
</div>


<main id="content" class="home__main">
    <section class="home__blocks">
        <div class="container">

            <div class="home__blocks-grid-wrap">
                <div class="home__blocks-grid">

                    <a href="https://www.economicsobservatory.com/what-do-we-know-about-labour-market-power-in-the-uk"
                        title="What do we know about labour market power in the UK?" style="color: #f88461;"
                        class="home__blocks-item home__blocks-item--padding home__blocks-item--style1 wow fadeInDown"
                        data-wow-offset="80" data-wow-duration="1s" data-wow-delay="0.2s">
                        <div>
                            <div class="home__blocks-item-image" style="color:#f88461;">


                                <div class="image "
                                    style="border-color:#f88461;background-image: url('https://www.economicsobservatory.com/wp-content/uploads/2024/02/iStock-
                                </div>
                            </div>
                        </div>


                        <div>
                            <div class="home__blocks-item-cat" style="color:#f88461;">
                                <span>New Post</span> •
                            </div>

                            <h3 class="home__blocks-item-title">What do we know about labour market power in the UK?
                            </h3>

                            <div class="home__blocks-item-teaser display">
                                <p>The ability of powerful firms to set pay and other working conditions is a
                                    growing concern for economic researchers and policy-makers. Much of the evidence
                                    is from the United States, where labour market institutions and policies are
                                    very different. Hence the need for emerging UK-specific analysis.</p>
                            </div>
                        </div>
                    </a>

                    <a href="data-hub" class="home__events home__blocks-item--style6">

                        <div id="circles3"></div>

                        <script>
                            var animationData = { "v": "5.10.0", "fr": 30, "ip": 0, "op": 363, "w": 795, "h": 275, "nm": "Adatation_23", "ddd": 0, "assets": [{
                            var params = {
                                container: document.getElementById('circles3'),
                                renderer: 'svg',
                                loop: true,
```



Questions and answers about the economy.

NEW POST •

What do we know about labour market power in the UK?

The ability of powerful firms to set pay and other working conditions is a growing concern for economic researchers and policy-makers. Much of the evidence is from the United States, where labour market institutions and policies are very different. Hence the need for emerging UK-specific analysis.

SCHOOLS, UNIVERSITIES & TRAINING

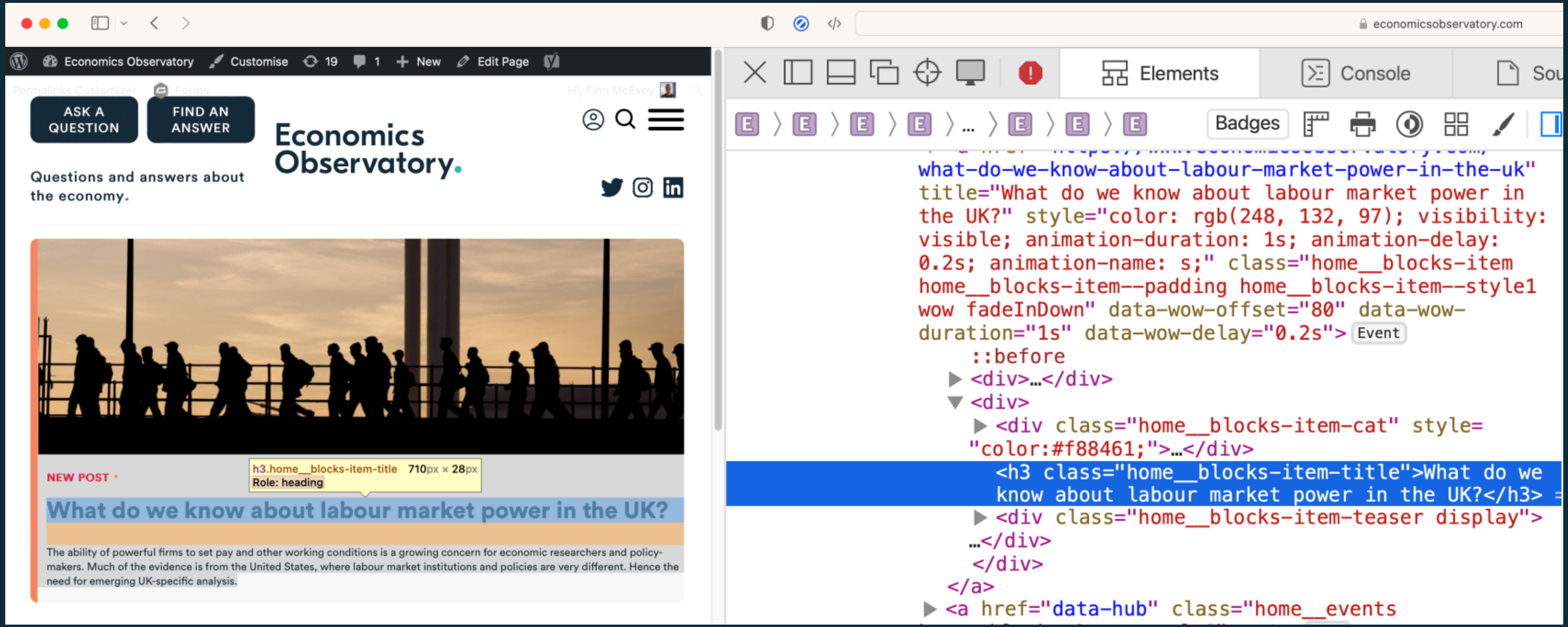
How can we reduce gender gaps in mathematics education?

LESSONS FROM HISTORY

How have minorities been treated by the UK's judicial system?

Scraping HTML Source.

We determine how the target data is defined using ‘inspect-element’



ASK A QUESTION

FIND AN ANSWER

Economics Observatory.

Questions and answers about the economy.

NEW POST •

h3.home__blocks-item-title 710px × 28px
Role: heading

What do we know about labour market power in the UK?

The ability of powerful firms to set pay and other working conditions is a growing concern for economic researchers and policy-makers. Much of the evidence is from the United States, where labour market institutions and policies are very different. Hence the need for emerging UK-specific analysis.

```
what-do-we-know-about-labour-market-power-in-the-uk"
title="What do we know about labour market power in
the UK?" style="color: rgb(248, 132, 97); visibility:
visible; animation-duration: 1s; animation-delay:
0.2s; animation-name: s;" class="home__blocks-item
home__blocks-item--padding home__blocks-item--style1
wow fadeInDown" data-wow-offset="80" data-wow-
duration="1s" data-wow-delay="0.2s"> Event
    ::before
  ▶ <div>…</div>
  ▼ <div>
    ▶ <div class="home__blocks-item-cat" style=
    "color:#f88461;">…</div>
      <h3 class="home__blocks-item-title">What do we
      know about labour market power in the UK?</h3>
    ▶ <div class="home__blocks-item-teaser display">
    …</div>
    </div>
  </a>
▶ <a href="data-hub" class="home__events
```

We see titles have a class “home_blocks-item-title”

Code-along.

A more advanced scraper

In this bonus practical session, we will use Google Colab to use scrape data from the Economics Observatory website using Python. Again, we can also embed a chart displaying the scraped data into your website, using VS Code and GitHub.

- Work through the following guided notebook: “Session_5_Scraping_advanced.ipynb” (open in Google Colab)

Learn more.

- In this session we have tried basic data scraping with Pandas and seen advanced scraping with BeautifulSoup
- There's still much more to learn
  - Choose your own projects (how can you make your job easier?)
  - Try bigger projects (scrape 100 pages, not just 1)
  - Try advanced tools (e.g. Selenium)

Learn more, responsibly.

- **Rate-limiting**: avoid making too many requests at once
- **Ethics**: Ensure your scraping activities do not harm the website's operation
- **Data Privacy**: Be mindful of personal data collection. Comply with relevant data protection laws (like GDPR).