資訊化社會科學夏季研習營

Summer Institute In Computational Social Science -Taiwan
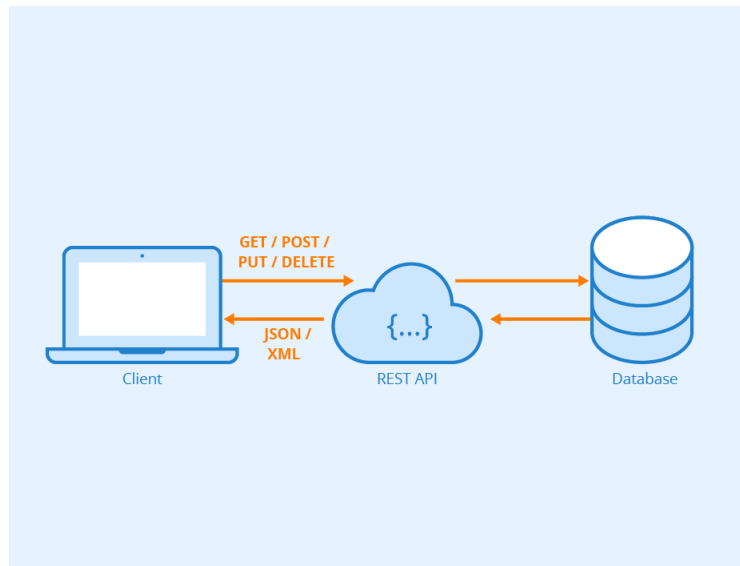
SICSS TAIWAN

Topic: Data Scraping – API
Instructor: Chung-pei Pien

International College of
INNOVATION
國立政治大學創新國際學院

Import Data: API

Import Data: API

There are rare opportunities to get csv or xlsx data.

Many data providers to avoid web scraping allow users to request Application Programming Interface (API) to get raw data.

There are four kinds of APIs

		R/Python package	
		Yes	No
API Key	Yes	US Census Bureau	--
	No	World Bank	Dcard

Import Data: API

How to Read and Understand API Documentations

https://betterprogramming.pub/how-to-read-and-understand-api-documentations-2f894b51d0b7
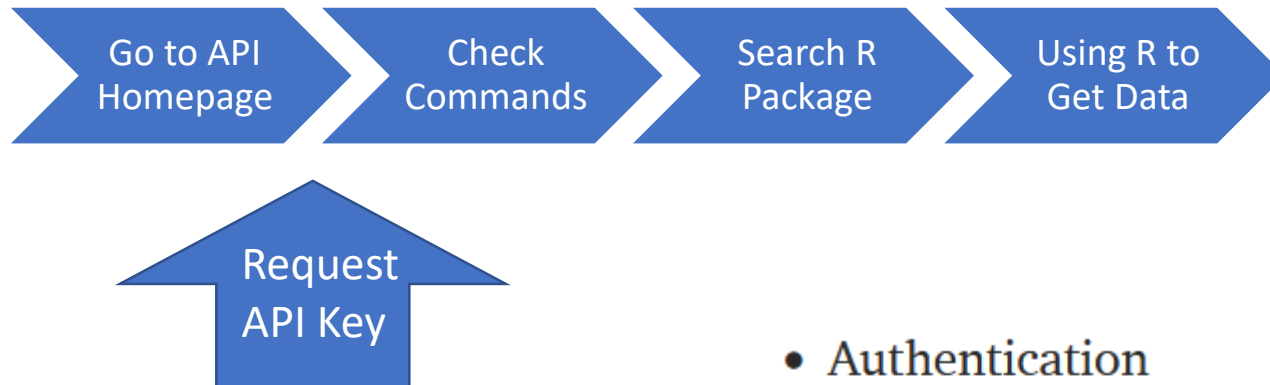
Import Data: API

Introduction

How to Read and Understand API Documentations

https://betterprogramming.pub/how-to-read-and-understand-api-documentations-2f894b51d0b7

Import Data: API

Dcard API

Let's go to Dcard API

| Go to API Homepage | Check Commands | Search R Package | Using R to Get Data |

Request API Key

- Authentication

- Object

- Request

- Response

Import Data: API

Dcard API

https://blog.jiatool.com/posts/dcard_api_v2/

Import Data: API

Dcard API

API 網址：**https://www.dcard.tw/service/api/v2**

說明	請求方法	路徑
全部文章	GET	/posts
看板資訊	GET	/forums
看板內文章列表	GET	/forums/{看板名稱}/posts
文章內文	GET	/posts/{文章ID}
文章內引用連結	GET	/posts/{文章ID}/links
文章內留言	GET	/posts/{文章ID}/comments

全部皆回傳 JSON 格式資料。

Import Data: API

library(jsonlite)

forum <- fromJSON("https://www.dcard.tw/service/api/v2/forums")

API 網址：**https://www.dcard.tw/service/api/v2**

說明	請求方法	路徑
全部文章	GET	/posts
看板資訊	GET	/forums
看板內文章列表	GET	/forums/{看板名稱}/posts
文章內文	GET	/posts/{文章ID}
文章內引用連結	GET	/posts/{文章ID}/links
文章內留言	GET	/posts/{文章ID}/comments

全部皆回傳 JSON 格式資料。

Import Data: API

Dcard API

library(jsonlite)

nccu <- fromJSON("https://www.dcard.tw/service/api/v2/forums/nccu /posts")

	id	alias	name	description
1	7f125e07-4460-4ea5-80b5-33f0e9aafa0c	midnightlab	午夜實驗室	午夜實驗室10/6、10/7即將在華山登場！這裏提供大家交流...
2	c0ed3f99-ed1c-49a8-b413-ed5e925aafe4	timecapsule	時光膠囊	Dcard 九週年最終章！告別動盪不安的 2020，說好了，2021...
3	cc2164ae-88cd-466d-b11b-19e4b1f6f78c	mother	母親節	Dcard 母親節限定板，告訴媽媽我愛妳！
4	11e8d941-0aad-4960-8348-b27c36270fd1	merryxmas	聖誕CiaoCiao	期間限定開啟！只要是和聖誕節有關的事情，或是和「禮物...
5	8d5cabde-d0df-479f-8ab7-153209317537	father	父親節	Dcard 父親節特輯，上傳爸爸帥照，寫下真摯文字，一起謝...
6	460a5276-0480-4af1-86df-14b3040cab08	5anniversary	五週年紀念	五週年徵文活動限定板，僅供活動徵文、討論，發文前請先...
7	4dd9f69b-9b33-4ba0-9cec-887d8c5091e4	bugreport	臨時回報	臨時回報版本問題
8	3c91f2b0-6374-4467-b557-2e5383680afb	ntu	臺灣大學	臺灣大學板，一個能讓你暢所欲言的地方。在這裡，卡友們...
9	caf9487a-1994-494d-81f2-7950438e7bef	nccu	政治大學	政治大學板，一個能讓你暢所欲言的地方。在這裡，卡友們...
10	204997a1-c63d-4dc4-965e-9f018e452d1d	nctu	交通大學	交通大學板，一個能讓你暢所欲言的地方。在這裡，卡友們...
11	5120400e-e5e6-4676-a1b8-75a13625ecea	tku	淡江大學	淡江大學板，一個能讓你暢所欲言的地方。在這裡，卡友們...
12	a8a2712f-80de-40c6-971f-c7bbb77716fe	ndhu	東華大學	東華大學板，一個能讓你暢所欲言的地方。在這裡，卡友們...
13	abc4fa45-a456-49c6-8948-8981ad0b0f97	ncku	成功大學	成功大學板，一個能讓你暢所欲言的地方。在這裡，卡友們...

library(jsonlite)

nccu <- fromJSON("https://www.dcard.tw/service/api/v2/forums/nccu/posts")

There are only 30 posts.

Import Data: API

Dcard API

```
For (i in 1:50) {

    Sys.sleep(45)

    last_post <- nccu$id[nrow(nccu)]

    before_nccu <-
    fromJSON(paste0("https://www.dcard.tw/service/api/v2/f
    orums/nccu/posts?limit=100", "&before=", last_post))

    nccu <- bind_rows(nccu, before_nccu)

}
```

Import Data: API

Dcard API Practice

分析近一個月以來Dcard上林智堅論文事件及安倍被刺殺的相關言論，請下載Dcard時事版的標題、內文及回應。

Let's go to World Bank API

Data

About the Indicators API Documentation

← Developer Information

About the Indicators API

Note: version 2 of the Indicators API was released several years ago, and version 1 was officially retired in November , 2018. Version 1 was discontinued entirely on June 19, 2020 (see link). If you access the older endpoints you will receive a "Resource not found" error. Use the information here to update your code.

The World Bank Indicators API provides access to nearly 16,000 time series indicators. Most of these indicators are available online through tools such as Databank and the Open Data website. The API provides programmatic access to this same data. Many data series date back over 50 years, and can be used to create interesting applications.

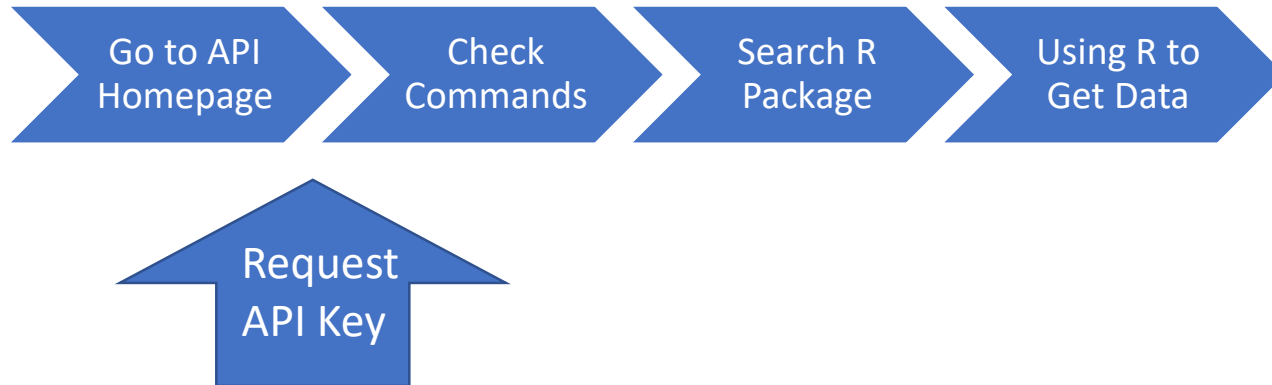The Indicators API provides access to over 45 databases, including:

- World Development Indicators
- International Debt Statistcs
- Doing Business
- Human Capital Index
- Subnational Poverty
- And many more

Import Data: API

World Bank API
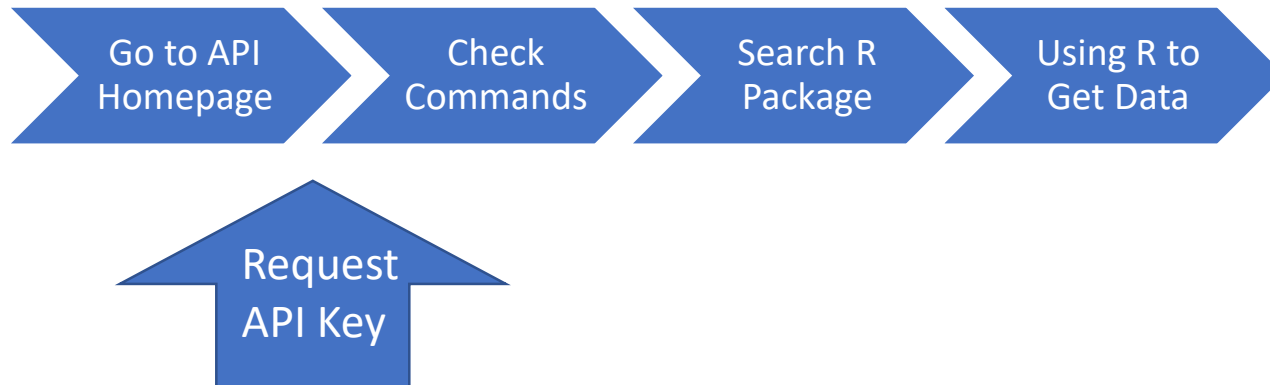
Let's go to World Bank API

Import Data: API

Let's go to World Bank API

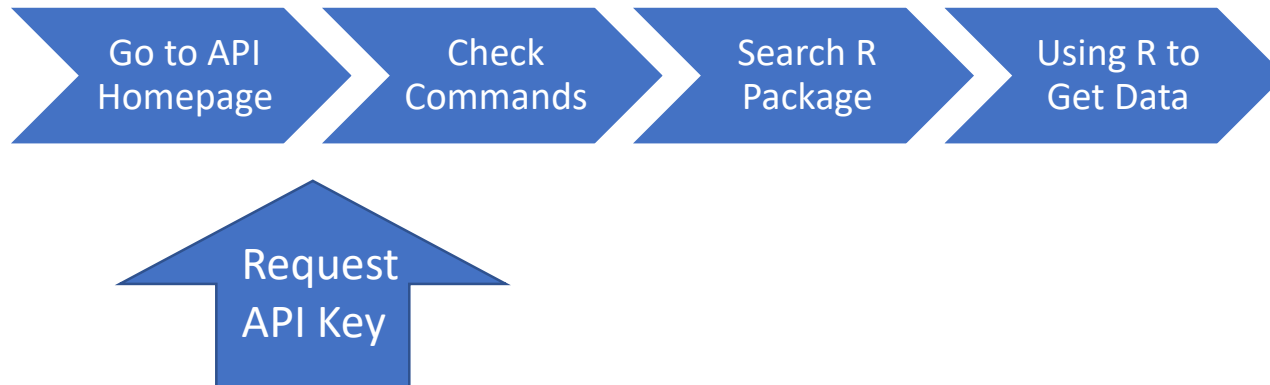| Go to API Homepage | Check Commands | Search R Package | Using R to Get Data |

Request API Key

API Access / Authentication

API keys and other authentication methods are no longer necessary to access the API.

World Bank API

Let's go to World Bank API

API doc:
https://datahelpdesk.worldbank.org/knowledgebase/articles/898581

Go to API Homepage → Check Commands → Search R Package → Using R to Get Data

Request API Key

Date and Date-Range: Date-range by year, month or quarter that scopes the result-set.

Examples:

- http://api.worldbank.org/v2/country/all/indicator/SP.POP.TOTL?date=2000
- http://api.worldbank.org/v2/country/chn;bra/indicator/DPANUSSPB?date=2012M01

Import Data: API

Let's go to World Bank API

| Go to API Homepage | Check Commands | Search R Package | Using R to Get Data |

Request API Key
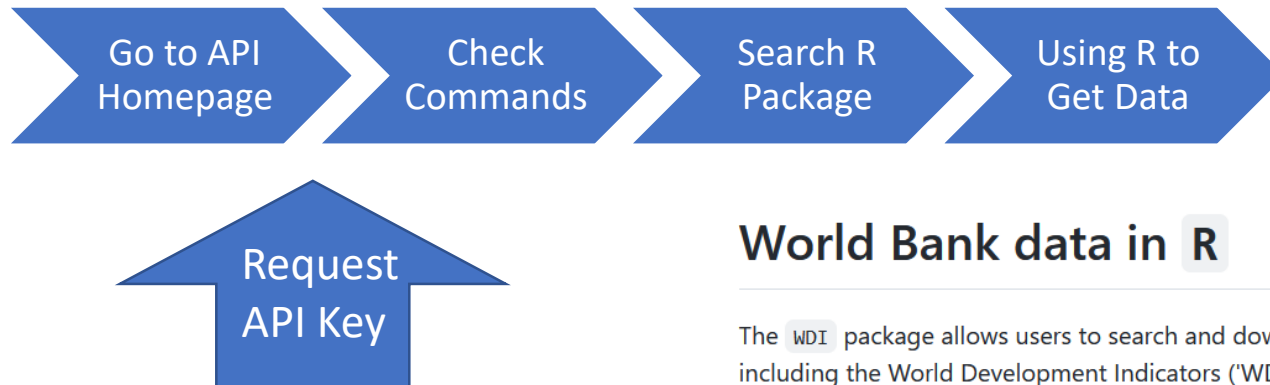
World Bank data in R

The `WDI` package allows users to search and downloa
including the World Development Indicators ('WDI'), I
Index, and Sub-national Poverty indicators.

`downloads 357K` `build failing` `R-CMD-check passing`

https://github.com/vincentarelbundock/WDI

`WDIsearch` uses grep and ignores cases, so you can also use regular expressions. For instance, if you are looking for GDP per capita in constant dollars:

```
WDIsearch('gdp.*capita.*constant')
     indicator          name
[1,] "GDPPCKD"          "GDP per Capita, constant US$, millions"
[2,] "NY.GDP.PCAP.KD"   "GDP per capita (constant 2000 US$)"
[3,] "NY.GDP.PCAP.KN"   "GDP per capita (constant LCU)"
[4,] "NY.GDP.PCAP.PP.KD" "GDP per capita, PPP (constant 2005 international $)"
```

Download and use the data

Download a series you like for the countries you like:

```
dat = WDI(indicator='NY.GDP.PCAP.KD', country=c('MX','CA','US'), start=1960, end=2012)
```

Look at the data:

```
install.packages("WDI")
library(WDI)

#search key words

ws_gini <- WDIsearch("gini")
```

Import Data: API

World Bank API

ws_gini <- WDIsearch("gini")

Which indictor we want?

	indicator	name
1	3.0.Gini	Gini Coefficient
2	3.0.Gini_nozero	Gini Coefficient (No Zero Income)
3	3.1.Gini	Gini, Rural
4	3.2.Gini	Gini, Urban
5	per_allsp_gini_rur	Gini inequality index reduction (%) - All Social Protection an...
6	per_allsp_gini_tot	Gini inequality index reduction (%) - All Social Protection an...
7	per_allsp_gini_urb	Gini inequality index reduction (%) - All Social Protection an...
8	per_lm_ac_gini_rur	Gini inequality index reduction (%) - Active Labor Market -r...

https://data.worldbank.org/

World Bank Open Data

Free and open access to global development data

Search data e.g. GDP, population, Indonesia

Browse by **Country** or **Indicator**

Gini index (World Bank estimate)

Income share held by highest 10%

https://data.worldbank.org/indicator/SI.POV.GINI?view=chart

Get the name from the link

gini <- WDI(indicator='SI.POV.GINI', start=1960, end=2012)

	iso2c	country	SI.POV.GINI Gini index (World Bank estimate)	year
1	ZH	Africa Eastern and Southern	NA	2012
2	ZH	Africa Eastern and Southern	NA	2011
3	ZH	Africa Eastern and Southern	NA	2010
4	ZH	Africa Eastern and Southern	NA	2009
5	ZH	Africa Eastern and Southern	NA	2008
6	ZH	Africa Eastern and Southern	NA	2007
7	ZH	Africa Eastern and Southern	NA	2006
8	ZH	Africa Eastern and Southern	NA	2005
9	ZH	Africa Eastern and Southern	NA	2004
10	ZH	Africa Eastern and Southern	NA	2003
11	ZH	Africa Eastern and Southern	NA	2002

22

Import Data: API

World Bank API

Specific countries' data

gini <- WDI(indicator = 'SI.POV.GINI', start = 1960, end =2012)

gini_country <- WDI(indicator = 'SI.POV.GINI', country = c('TR', 'TH'), start=1960, end=2012)

https://www.iban.com/country-codes

Thailand		TH	THA
Timor-Leste		TL	TLS
Togo		TG	TGO
Tokelau		TK	TKL
Tonga		TO	TON
Trinidad and Tobago		TT	TTO
Tunisia		TN	TUN
Turkey		TR	TUR
Turkmenistan		TM	TKM

23

Import Data: API

World Bank API Practice

The inflation has become a seious issue around the world.

Please use world bank's API to download all EU countries' annual inflation rates from 1960 to 2020.

Hint: How to collect all EU countries' Alpha-2 code and import them to R?

Import Data: API

US Census Bureau API

Let's go to US Census Bureau's Economic Indictors API.

https://www.census.gov/data/developers/data-sets.html

Import Data: API

US Census Bureau API

| Go to API Homepage | Check Commands | Search R Package | Using R to Get Data |

Request API Key

Click "Request a key"

Request A Key

Organization Name: NCCU ICI

Email Address: torrent@nccu.edu.tw

☑ I agree to the terms of service

Submit Key Request

Import Data: API

US Census Bureau API

Census Data API Service <no-reply@census.gov>   5:24 PM (0 minutes ago)
to torrent ▼

Hello!

Thank you for your interest in the Census Data API. Your API key is *fb56ed955*█████████████████ Please click here to activate your key.

Save this email for future reference.

Have Fun,

The Census Bureau API Team
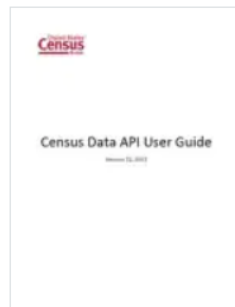
Follow @uscensusbureau on twitter for API updates.

Congratulations! Your key has been activated. You may now use it to query the Data API.

Happy querying!

Import Data: API

US Census Bureau API

https://www.census.gov/data/developers/guidance/api-user-guide.html

Census Data API User Guide

JANUARY 14, 2021

The purpose of this user guide is to instruct developers and researchers on how to use the Census Data Application Programming Interface (API) to request data from U.S. Census Bureau datasets.

Download Census Data API User Guide [PDF - <1.0 MB]

Overview	What is the API	Available Data	Core Concepts	Example API Queries	Help & Contact Us

Overview

Import Data: API

US Census Bureau API

01

install.packages("censusapi")
library(censusapi)

US Census Bureau provides thousands databases. The website of US Census Bureau just likes a maze.

Censusapi provides tools to get data.

Get Database List	Get a Database's Variable List	Get the data you want

29

Import Data: API

US Census Bureau API

#Get Database list

apis <- listCensusApis()

	title	name	vintage	url
821	Economic Surveys: Annual Business Survey: Annual Business...	abscb	2017	http://api.census.gov/data/2017/abscb
986	Economic Surveys: Annual Business Survey: Annual Business...	abscb	2018	http://api.census.gov/data/2018/abscb
1131	Economic Surveys: Annual Business Survey: Annual Business...	abscb	2019	http://api.census.gov/data/2019/abscb
827	Economic Surveys: Annual Business Survey: Annual Business...	abscbo	2017	http://api.census.gov/data/2017/abscbo
987	Economic Surveys: Annual Business Survey: Annual Business...	abscbo	2018	http://api.census.gov/data/2018/abscbo
1132	Economic Surveys: Annual Business Survey: Annual Business...	abscbo	2019	http://api.census.gov/data/2019/abscbo
829	Annual Business Survey: Company Summary: 2017	abscs	2017	http://api.census.gov/data/2017/abscs
988	Economic Surveys: Annual Business Survey: Annual Business...	abscs	2018	http://api.census.gov/data/2018/abscs
1133	Economic Surveys: Annual Business Survey: Annual Business...	abscs	2019	http://api.census.gov/data/2019/abscs
1139	Economic Surveys: Annual Business Survey: Annual Business...	absnesd	2018	http://api.census.gov/data/2018/absnesd
1140	Economic Surveys: Annual Business Survey: Annual Business...	absnesdo	2018	http://api.census.gov/data/2018/absnesdo
1028	Annual Economic Surveys: Annual Business Survey	abstcb	2018	http://api.census.gov/data/2018/abstcb
751	American Community Survey: 1-Year Estimates: Detailed Tab...	acs/acs1	2005	http://api.census.gov/data/2005/acs/acs1
756	American Community Survey: 1-Year Estimates: Detailed Tab...	acs/acs1	2006	http://api.census.gov/data/2006/acs/acs1

Import Data: API

US Census Bureau API

#Get a database's variable list

vars_list <- listCensusMetadata(
  name = "acs/acs5",
  vintage = 2020,
  type = "variables")

	name	label	concept	p
1	B24022_060E	Estimate!!Total!!Female!!Service occupations!!Food preparati...	SEX BY OCCUPATION AND MEDIAN EARNINGS IN THE PAST...	in
2	B19001B_014E	Estimate!!Total!!$100,000 to $124,999	HOUSEHOLD INCOME IN THE PAST 12 MONTHS (IN 2017 I...	in
3	B07007PR_019E	Estimate!!Total!!Moved from different municipio!!Foreign bo...	GEOGRAPHICAL MOBILITY IN THE PAST YEAR BY CITIZENSH...	in
4	B19101A_004E	Estimate!!Total!!$15,000 to $19,999	FAMILY INCOME IN THE PAST 12 MONTHS (IN 2017 INFLATI...	in
5	B24022_061E	Estimate!!Total!!Female!!Service occupations!!Building and g...	SEX BY OCCUPATION AND MEDIAN EARNINGS IN THE PAST...	in
6	B19001B_013E	Estimate!!Total!!$75,000 to $99,999	HOUSEHOLD INCOME IN THE PAST 12 MONTHS (IN 2017 I...	in
7	B07007PR_018E	Estimate!!Total!!Moved from different municipio!!Foreign bo...	GEOGRAPHICAL MOBILITY IN THE PAST YEAR BY CITIZENSH...	in
8	B19101A_005E	Estimate!!Total!!$20,000 to $24,999	FAMILY INCOME IN THE PAST 12 MONTHS (IN 2017 INFLATI...	in
9	B19001B_012E	Estimate!!Total!!$60,000 to $74,999	HOUSEHOLD INCOME IN THE PAST 12 MONTHS (IN 2017 I...	in
10	B24022_062E	Estimate!!Total!!Female!!Service occupations!!Personal care ...	SEX BY OCCUPATION AND MEDIAN EARNINGS IN THE PAST...	in
11	B01001B_029E	Estimate!!Total!!Female!!65 to 74 years	SEX BY AGE (BLACK OR AFRICAN AMERICAN ALONE)	in

Import Data: API

US Census Bureau API

If we want all California counties':

1. SEX BY AGE
2. MEDIAN AGE BY SEX
3. Median household income in the past 12 months

The names of variables are:

1. B01001_001E
2. B01002_001E
3. B19013_001E

	name	label
1	B24022_060E	Estimate!!Total!!Female!!Service occupa
2	B19001B_014E	Estimate!!Total!!$100,000 to $124,999
3	B07007PR_019E	Estimate!!Total!!Moved from different n
4	B19101A_004E	Estimate!!Total!!$15,000 to $19,999
5	B24022_061E	Estimate!!Total!!Female!!Service occupa
6	B19001B_013E	Estimate!!Total!!$75,000 to $99,999
7	B07007PR_018E	Estimate!!Total!!Moved from different n
8	B19101A_005E	Estimate!!Total!!$20,000 to $24,999
9	B19001B_012E	Estimate!!Total!!$60,000 to $74,999
10	B24022_062E	Estimate!!Total!!Female!!Service occupa
11	B01001B_029E	Estimate!!Total!!Female!!65 to 74 years
12	B20005A_021E	Estimate!!Total!!Male!!Worked full-time

```r
asc_df <- getCensus(name = "acs/acs5", vintage = 2020,
                    key = "fb56e                      ",
                    vars = c("B01001_001E", "NAME", "B01002_001E", "B19013_001E", "GEO_ID"),
                    region = "county:*", regionin = "state:06")
```

Import Data: API

01

Merge SEX BY AGE, MEDIAN AGE BY SEX , Median household income into US COVID-19 Vaccination data to evaluate the above population variables' impacts on change in vaccination rate in Biden's first-year presidency.

COVID-19 Vaccinations in the United States,County Vaccinations

View Data | Visualize ∨ | Export | API | ···

Overall US COVID-19 Vaccine administration and vaccine equity data at county level. Data represents all vaccine partners including jurisdictional partner clinics, retail pharmacies, long-term care facilities, dialysis centers, Federal Emergency Management Agency and Health Resources and Services Administration partner sites, and federal entity facilities.

Updated
July 16, 2022

Data Provided by
IISInfo

Featured Content Using this Data

https://data.cdc.gov/Vaccinations/COVID-19-Vaccinations-in-the-United-States-County/8kkx-amqh

Import Data: API

```
#Read vaccination rate county level csv file
vac <- read.csv("COVID-19_Vaccinations_in_the_United_States_County.csv")

#Select the data from 01/20/2021 and 01/19/2022
vac <- vac %>%
  filter(Date %in% c("01/20/2021", "01/19/2022")) %>%
  select(1:6)

#Transform long table to wide table
vac_sp <- vac[,-3] %>%
  spread(Date, Completeness_pct)

#Calculate the change from 01/20/2021 to 01/19/2022
vac_sp <- vac_sp %>%
  mutate(one_year = `01/19/2022` - `01/20/2021`)
```

```
asc_df <- getCensus(name = "acs/acs5", vintage = 2020,
                    key = "fb56ed
                    vars = c("B01001_001E", "NAME", "B01002_001E", "B19013_001E", "GEO_ID"),
                    region = "county:*")
```

#Create the county code column
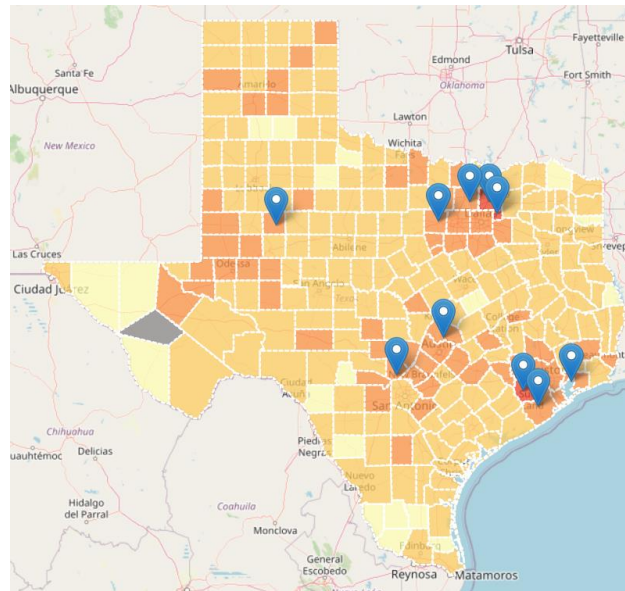asc_df$FIPS <- paste0(asc_df$state, asc_df$county)

asc_df <- asc_df %>%
  left_join(vac_sp, by = "FIPS")

Import Data: API

US Census Bureau Practice

Please download all US counties' 2020 unemployment rate and merget it into US 2020 presidential election result (2020_US_County_Level_Presidential_Results.csv).

Hint: Download acs_researchers_handbook_2020_ch05.pdf to figure out how to get *unemployment rate* in US Census Bureau API.

Import Data: API

Create a Texas income leaflet map

1. Use censusapi to collect Texas counties' 2022 income median data
2. Merge the income data into TX shapefile
3. Create a leaflet map to show all counties' income median in Texas and put blue balloons on the top 10 income countries.

Data Project