How to retrieve data with the loc[] method

What you'll learn

- How to use `loc[]` to retrieve data in DataFrames

- Select rows and columns based on *label*
  - Selecting specific cells of data
  - Selecting single rows
  - Selecting single columns

- How to select "slices" of data
  - select multiple rows or columns at the same time

Quick review of dataframe structure

HERE'S THE DATA WE'LL BE USING

- We'll use the following data for the examples in this presentation

- This is actually a subset of a more complete dataset you'll work with elsewhere

country_code	country	continent	population	gdp
USA	United States	North America	323,405,935	18,624,475,000,000
CHN	China	Asia	1,378,665,000	11,190,992,550,230
JPN	Japan	Asia	126,994,511	4,949,273,341,994
DEU	Germany	Europe	82,348,669	3,477,796,274,497
GBR	United Kingdom	Europe	65,595,565	2,650,850,178,102
FRA	France	Europe	66,859,768	2,465,134,297,439

EACH ROW AND EACH COLUMN HAVE AN INTEGER INDEX

You learned about this in the lesson on `iloc[]`

column index

	0	1	2	3	4
	country_code	**country**	**continent**	**population**	**gdp**
0	USA	United States	North America	323,405,935	18,624,475,000,000
1	CHN	China	Asia	1,378,665,000	11,190,992,550,230
2	JPN	Japan	Asia	126,994,511	4,949,273,341,994
3	DEU	Germany	Europe	82,348,669	3,477,796,274,497
4	GBR	United Kingdom	Europe	65,595,565	2,650,850,178,102
5	FRA	France	Europe	66,859,768	2,465,134,297,439

row index

ROWS AND COLUMNS CAN ALSO BE REFERENCED BY "LABEL"

column label

country_code	country	continent	population	gdp
USA	United States	North America	323,405,935	18,624,475,000,000
CHN	China	Asia	1,378,665,000	11,190,992,550,230
JPN	Japan	Asia	126,994,511	4,949,273,341,994
DEU	Germany	Europe	82,348,669	3,477,796,274,497
GBR	United Kingdom	Europe	65,595,565	2,650,850,178,102
FRA	France	Europe	66,859,768	2,465,134,297,439

row label

METHODS FOR SELECTING DATA

THERE ARE TWO PANDAS METHODS FOR SELECTING ROWS AND COLUMNS

- The `loc[]` method selects by label

- The `iloc[]` method selects by integer index

- We can use both of these to select:
  - single rows of data
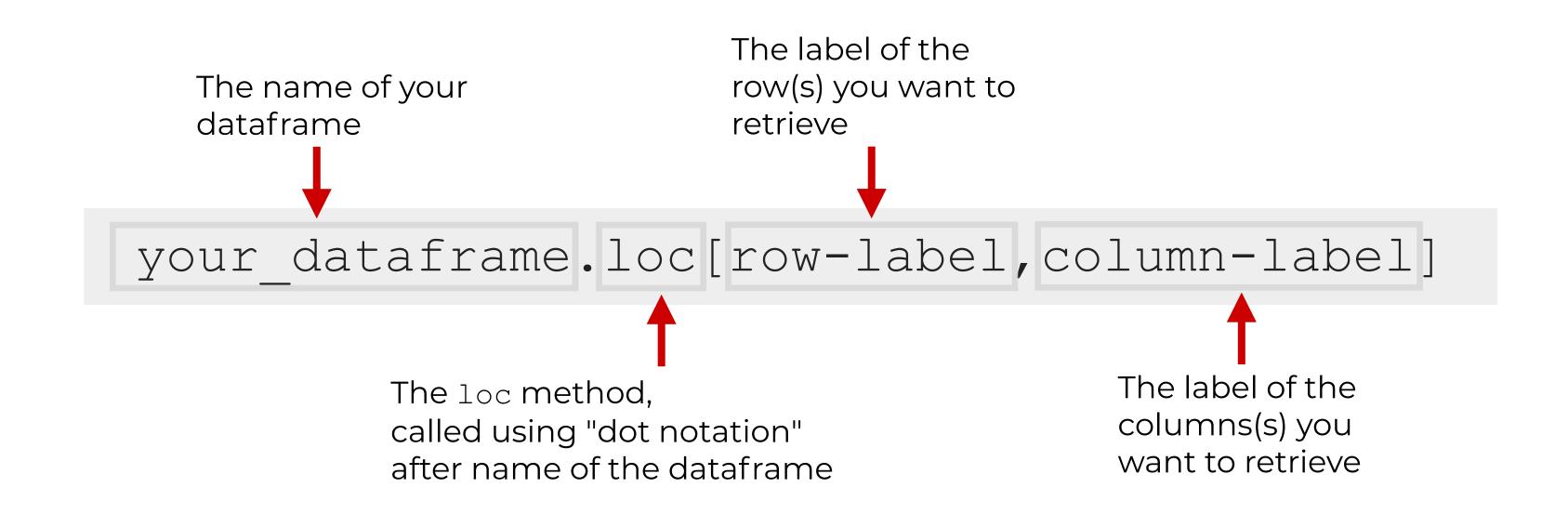 - slices of data

Introduction to the pandas loc method

THE LOC METHOD ALLOWS YOU TO LOCATE DATA BY LABEL
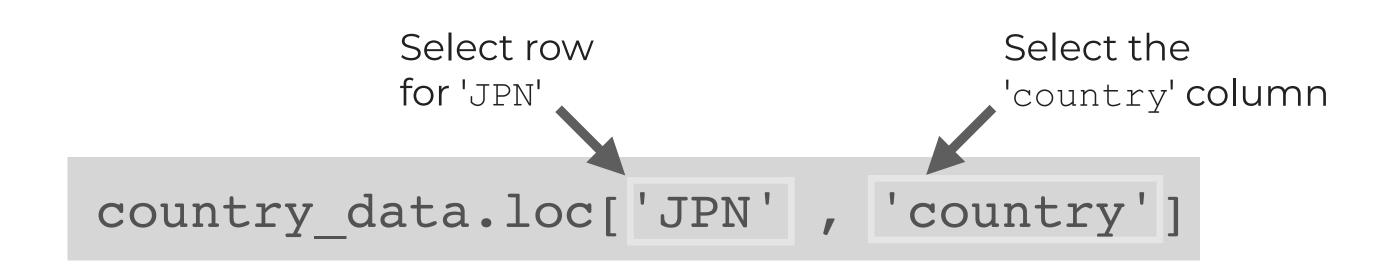
Here, we're selecting a column by "label"

country_code	country	continent	population	gdp
USA	United States	North America	323,405,935	18,624,475,000,000
CHN	China	Asia	1,378,665,000	11,190,992,550,230
JPN	Japan	Asia	126,994,511	4,949,273,341,994
DEU	Germany	Europe	82,348,669	3,477,796,274,497
GBR	United Kingdom	Europe	65,595,565	2,650,850,178,102
FRA	France	Europe	66,859,768	2,465,134,297,439

SYNTAX OF THE LOC[] METHOD

The name of your
dataframe

The label of the
row(s) you want to
retrieve

`your_dataframe.loc[row-label,column-label]`

The `loc` method,
called using "dot notation"
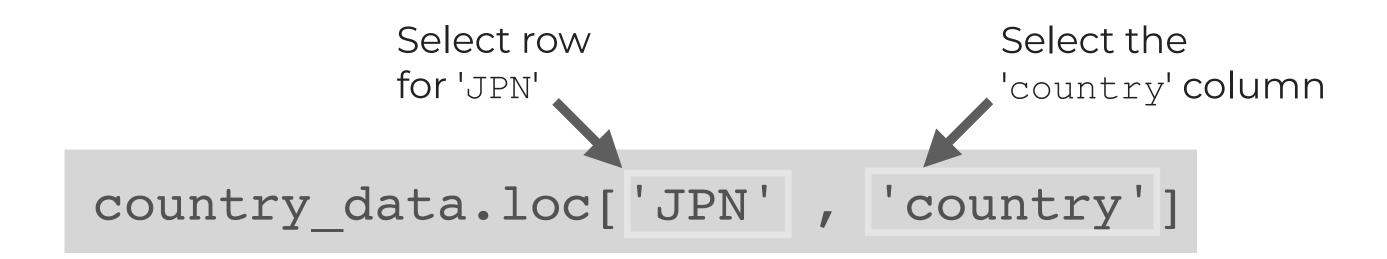after name of the dataframe

The label of the
columns(s) you
want to retrieve

Example: how to select a single cell with loc

To select a cell of data, we need to specify the row and column with loc

Select row
for `'JPN'`

Select the
`'country'` column

`country_data.loc[`**`'JPN'`**`,` **`'country'`**`]`

SELECT A CELL WITH LOC

Select row
for 'JPN'

Select the
'country' column

```
country_data.loc['JPN' , 'country']
```

country_code	country	continent	population	gdp
USA	United States	North America	323,405,935	18,624,475,000,000
CHN	China	Asia	1,378,665,000	11,190,992,550,230
JPN	Japan	Asia	126,994,511	4,949,273,341,994
DEU	Germany	Europe	82,348,669	3,477,796,274,497
GBR	United Kingdom	Europe	65,595,565	2,650,850,178,102
FRA	France	Europe	66,859,768	2,465,134,297,439

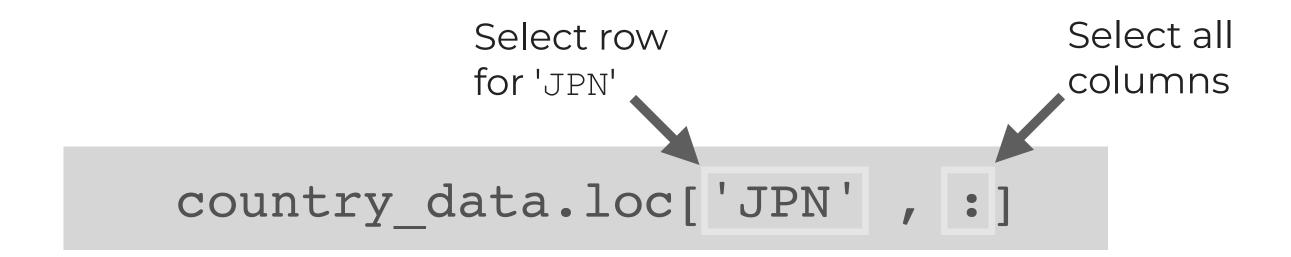# How to select a single row with loc

Syntax to select a row of data with loc[]

A row label indicates that we want to retrieve the data for a particular row

```
your_dataframe.loc[row-label, : ]
```

The ':' indicates that we want to retrieve all columns

SELECT A ROW WITH LOC

Select row
for 'JPN'

Select all
columns

```
country_data.loc['JPN' , :]
```

country_code	country	continent	population	gdp
USA	United States	North America	323,405,935	18,624,475,000,000
CHN	China	Asia	1,378,665,000	11,190,992,550,230
JPN	Japan	Asia	126,994,511	4,949,273,341,994
DEU	Germany	Europe	82,348,669	3,477,796,274,497
GBR	United Kingdom	Europe	65,595,565	2,650,850,178,102
FRA	France	Europe	66,859,768	2,465,134,297,439

How to select a single column with loc

SYNTAX TO SELECT A COLUMN OF DATA WITH LOC[]
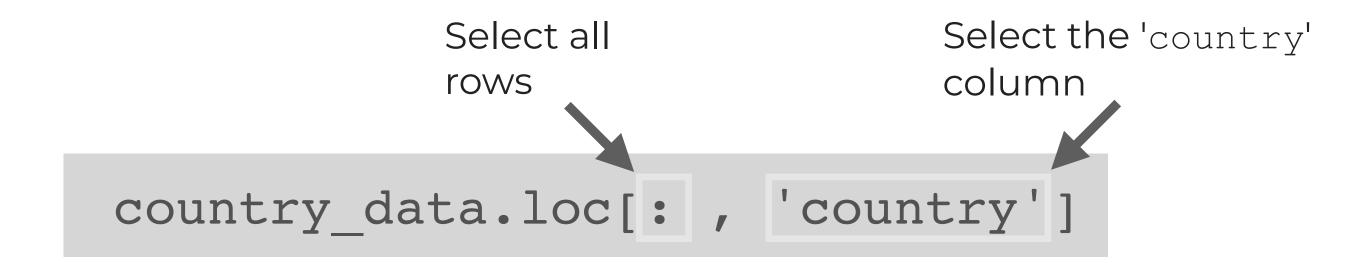
The ':' indicates that we want
to retrieve all rows

```
your_dataframe.loc[: , column-label ]
```

The column label index
indicates that we want to
retrieve the data for a
particular column

SELECT A COLUMN WITH LOC
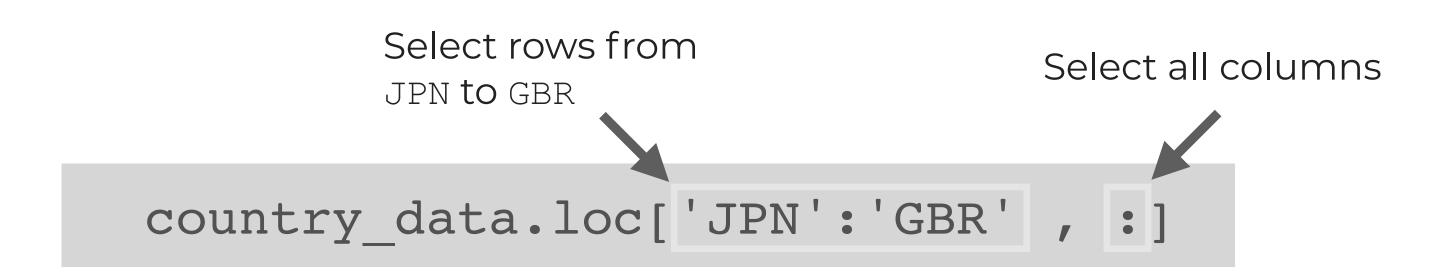
Select all
rows

Select the 'country'
column

country_data.loc[: , 'country']

country_code	country	continent	population	gdp
USA	United States	North America	323,405,935	18,624,475,000,000
CHN	China	Asia	1,378,665,000	11,190,992,550,230
JPN	Japan	Asia	126,994,511	4,949,273,341,994
DEU	Germany	Europe	82,348,669	3,477,796,274,497
GBR	United Kingdom	Europe	65,595,565	2,650,850,178,102
FRA	France	Europe	66,859,768	2,465,134,297,439

Slicing rows *and* columns

SYNTAX TO SELECT MULTIPLE ROWS OF DATA WITH LOC[]

The label of the first row
you want to retrieve

```
your_dataframe.loc[ start:stop, :]
```

The label of the row that
we want to "stop" at

(this row _will_ be retrieved)

SELECT MULTIPLE ROWS

Select rows from
JPN to GBR

Select all columns

`country_data.loc[`**`'JPN':'GBR'`**` , `**`:`**`]`

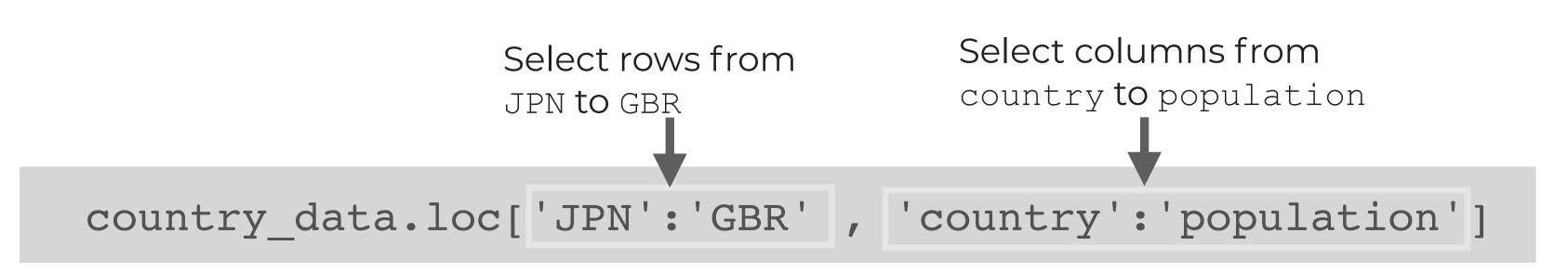| country_code	country	continent	population	gdp
USA	United States	North America	323,405,935	18,624,475,000,000
CHN	China	Asia	1,378,665,000	11,190,992,550,230
JPN	Japan	Asia	126,994,511	4,949,273,341,994
DEU	Germany	Europe	82,348,669	3,477,796,274,497
GBR	United Kingdom	Europe	65,595,565	2,650,850,178,102
FRA	France	Europe	66,859,768	2,465,134,297,439

SELECT MULTIPLE COLUMNS

Select all rows

Select columns from country to population

```
country_data.loc[: , 'country':'population']
```

country_code	country	continent	population	gdp
USA	United States	North America	323,405,935	18,624,475,000,000
CHN	China	Asia	1,378,665,000	11,190,992,550,230
JPN	Japan	Asia	126,994,511	4,949,273,341,994
DEU	Germany	Europe	82,348,669	3,477,796,274,497
GBR	United Kingdom	Europe	65,595,565	2,650,850,178,102
FRA	France	Europe	66,859,768	2,465,134,297,439

SELECT SUBSET OF CELLS

Select rows from
JPN **to** GBR

Select columns from
country **to** population

```
country_data.loc['JPN':'GBR' , 'country':'population']
```

country_code	country	continent	population	gdp
USA	United States	North America	323,405,935	18,624,475,000,000
CHN	China	Asia	1,378,665,000	11,190,992,550,230
JPN	Japan	Asia	126,994,511	4,949,273,341,994
DEU	Germany	Europe	82,348,669	3,477,796,274,497
GBR	United Kingdom	Europe	65,595,565	2,650,850,178,102
FRA	France	Europe	66,859,768	2,465,134,297,439

Recap

RECAP OF WHAT WE LEARNED

- How to retrieve data using `loc[]`

    - select single rows by label

    - select single columns, by column name

- How to select *slices* of data

    - multiple rows

    - multiple columns

    - subset of cells