

Table of Contents

- [1 IO: Files](#)
- [2 Write data into a file](#)
- [3 Close the file after use](#)
- [4 Dealing with errors](#)
- [5 Reading CSV files](#)
- [6 Read input from the keyboard](#)

1 IO: Files

- **File:** a *sequence* of data held in a *storage medium*, that can be either *volatile* or *not*
 - User data are written on the file according to a *custom organization* that reflects user's needs
 - **Data structure:** a collection of data elements organized in some way (e.g., list, dictionary, set)

A file is a custom, *permanent* data structure to hold data

- A file data structure can be used to hold and represent virtually anything
 - Image, in different formats, such as jpeg, png, svg, gif ...
 - Data of interest: Genome maps, financial data, climate data, traffic logs, medical data....
 - Log about a running program, such as a Web server, or your own python program ...
 - Data collected by sensors during experiments
 - Text of a novel, a poem, a song, ...
 - Specification of a web page, in HTML, CSS, JavaScript, ...
 - Notes about ... anything!
 -

In [48]:

```
1 %ls
```

```
110-S21-19-Jupyter_notebooks.ipynb  TurtleCode/
110-S21-19-Jupyter_notebooks.pdf      Untitled.ipynb
IO_Files.ipynb                        accounts.txt
L0.py*                                cmu_logo.png
L1.py*                                data.txt
L10.py*                               data.txt~
L11.py*                               data2.txt
L12.py*                               employee_addresses.csv
L14.py*                               frog.txt
L15.py*                               hw05.py*
L17.py*                               price_list.csv
L1_1.py*                              scratch.py*
L8.py*                                tales_two_cities.txt
L9.py*                                untitled3.py*
Lab-1.py                              untitled4.py*
Mall_Customers.csv
```

In [49]:

```
1 %cat tales_two_cities.txt
```

CHAPTER I.
The Period

It was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the epoch of belief, it was the epoch of incredulity, it was the season of Light, it was the season of Darkness, it was the spring of hope, it was the winter of despair, we had everything before us, we had nothing before us, we were all going direct to Heaven, we were all going direct the other way--in short, the period was so far like the present period, that some of its noisiest authorities insisted on its being received, for good or for evil, in the superlative degree of comparison only.

CHAPTER I. The Period

It was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the epoch of belief, it was the epoch of incredulity, it was the season of Light, it was the season of Darkness, it was the spring of hope, it was the winter of despair, we had everything before us, we had nothing before us, we were all going direct to Heaven, we were all going direct the other way--in short, the period was so far like the present period, that some of its noisiest authorities insisted on its being received, for good or for evil, in the superlative degree of comparison only.

In []:

```
1 %cat IO_Files.ipynb
```

...

Text and Binary Files

Files can be broadly classified as:

▪ Binary

- *Images*: jpg, png, gif, bmp, tiff, psd, ...
- *Videos*: mp4, mkv, avi, mov, mpg, vob, ...
- *Audio*: mp3, aac, wav, flac, mka, wma, ...
- *Documents*: pdf, doc, xls, ppt, docx, odt, ...
- *Archive*: zip, rar, 7z, tar, iso, ...
- *Database*: mdb, accde, frm, sqlite, ...
- *Executable*: exe, dll, so, class, ...

▪ Text

- *Web tools*: html, xml, css, svg, json, ...
- *Source code*: c, cpp, h, cs, js, py, java, php, sh, ...
- *Documents*: txt, tex, markdown, asciidoc, rtf, ps, ...
- *Configuration*: ini, cfg, rc, reg, ...
- *Tabular data*: csv, tsv, ...

Text files: records and fields

▪ Text:

- Human-readable: mostly composed of *printable* characters
- Can be read / write using any text editor / viewer program
- Organized in **multiple records** separated by **newline characters**
- Each record is a piece of information, possibly structured in multiple **fields**

- Six records
- Four fields (at most):
First name, Last name, ID, Sex

John Smith	642876	M
Adam Smith	787294	M
Ann White.	889220	F
Joan Black	627291	F
Mary Brown	78979	

In [54]:

```
1 %cat accounts.txt
```

```
John Smith      642876 M
Adam Smith     787294 M
Ann White.      889220 F
Joan Black      627291 F
Mary Brown     78979 F
```

- Three records
- Variable number of fields per record

```
Old pond
Frog jumps in
Sound of water
```

In [9]:

```
1 %cat frog.txt
```

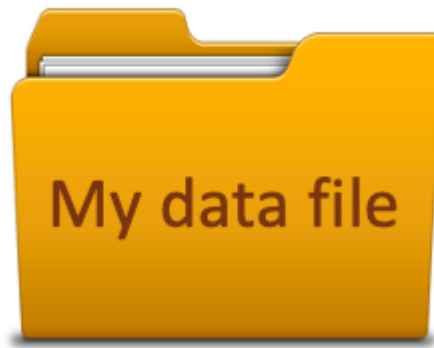
```
Old pond
Frog jumps in
Sound of water
```

- Four records
- Nine fields (at most), specified in the first record

```
STATION,STATION_NAME,ELEVATION,LATITUDE,LONGITUDE,DATE,HPCP,Measurement Flag,Quality Flag
COOP:310301,ASHEVILLE NC US,682.1,35.5954,-82.5568,20100101 00:00,99999,],
COOP:310301,ASHEVILLE NC US,682.1,35.5954,-82.5568,20100101 01:00,0,g,
COOP:310301,ASHEVILLE NC US,682.1,35.5954,-82.5568,20100102 06:00,1, ,
```

What type of **operations** can we do on / using a file (i.e., data storage)?

Data storage: storing data in a named location (a 'known' place, a **file**) that can be accessed (**open**) later for **read / write / update operations**, and can put aside (**closed**) when done, and it can also be **removed** if stored data are not anymore needed



Create a named file

✓ **Open** a *new* file



✓ **Open** an *existing* file

✓ **Read** *existing* data

✓ **Write** *new* data

✓ **Modify / write** *existing* data



Temporary shutdown an existing file

✓ **Close** *existing* file



Delete an existing file from the system

✓ **Remove** an *existing* file

Open a file (and let's focus on text files): open () function

- ✓ While each OS has its own way to deal with file (**file system** of an OS), Python achieves **OS-independence** by accessing a file through a **file handle** that holds the reference to the file in the system

- **Open** a file to make I/O on it, using the function:

```
file_handle = open(file_name, <modes>)
```

↑ ↑ ↑
TextIOWrapper str str
object type (stream) object type object type

```
f = open('data.txt', 'rt')  
f = open('data.txt', 'r')  
f = open('data.txt', 'w')  
f = open('data.txt', 'r+')  
f = open('data.txt', 'w+')
```

modes:

1. Read/Write
2. Text/Binary

'r' open an existing file for reading (default)
'w' open for writing, *truncating* file to 0 bytes if it exists, or creating a new file otherwise
'x' create a new file and open it for writing
'a' open for writing, *appending* to end of file if it exists, or creating a new file otherwise
'+' open a file for updating (both reading and writing)
'b' binary mode
't' text mode (default)

Default: 'rt'

```
In [55]: 1 %ls *.txt *.csv
```

```
Mall_Customers.csv      data2.txt              new.txt  
accounts.txt            employee_addresses.csv   price_list.csv  
data.txt                frog.txt                tales_two_cities.txt
```

```
In [59]: 1 f = open('data.txt', 'tr')
```

```
In [60]: 1 f
```

```
Out[60]: <_io.TextIOWrapper name='data.txt' mode='tr' encoding='UTF-8'>
```

The 'tr' flag doesn't need to be there!

```
In [61]: 1 f = open('data.txt')
```

```
In [62]: 1 f
```

```
Out[62]: <_io.TextIOWrapper name='data.txt' mode='r' encoding='UTF-8'>
```

```
In [66]: 1 %pwd
```

```
Out[66]: '/Users/giannidicaro/Dropbox/CMU-Lectures/PrinciplesOfComputing-15110/Code'
```

A **full path** (in the local file tree) can be specified to find a file

```
In [67]: 1 f = open('/Users/giannidicaro/workplan.tex')
```

Read from a file: `read()` method

- **Read** data from a file open with the `r` or `r+` mode flag (or, also, `w+`, `a+`)
`string_with_data = file_handle.read(number_of_bytes)`

One byte == One character

The method `read()` reads at most `number_of_bytes` bytes from the file, from **current position** in file

```
In [20]: 1 %cat data.txt
```

```
This line is 26 characters
Line 1: 0.1 5.4 2 4 20 .03
Line 2: 1 3. 2.0 43 12
Line 3: 1 2
This is the last record
New line: 0 3 5.5
```



```
In [68]: 1 f = open('data.txt', 'tr')
         2 s = f.read(10)
```

```
In [69]: 1 s
```

```
Out[69]: 'This line '
```

```
In [17]: 1 print(s)
```

This line

```
In [21]: 1 s = f.read(8)
```

```
In [22]: 1 s
```

```
Out[22]: 'is 26 ch'
```

How to go *back* to a certain position?

`.seek(pos)`

Go back to the **beginning of the file?**

```
In [24]: 1 f.seek(0)
```

```
Out[24]: 0
```

```
In [25]: 1 f.read(20)
```

```
Out[25]: 'This line is 26 char'
```

```
In [27]: 1 f.read(1000)
```

```
Out[27]: 'acters\nLine 1: 0.1 5.4 2 4 20 .03\nLine 2: 1 3. 2.0 43 12 \nLine 3:
1 2 \nThis is the last record\nNew line: 0 3 5.5\n'
```

To read the **whole file** in a shot

```
In [30]: 1 f.seek(0)
         2 s = f.read()
```

```
In [31]: 1 print(s)
```

```
This line is 26 characters
Line 1: 0.1 5.4 2 4 20 .03
Line 2: 1 3. 2.0 43 12
Line 3: 1 2
This is the last record
New line: 0 3 5.5
```

Big MORAL: strings data types are important because we read everything into a string!

Read a file **line by line, record by record?**

```
In [33]: 1 f.seek(0)
         2
         3 for record in f:
         4     print(record, end = '')
```

```
This line is 26 characters
Line 1: 0.1 5.4 2 4 20 .03
Line 2: 1 3. 2.0 43 12
Line 3: 1 2
This is the last record
New line: 0 3 5.5
```

```
In [34]: 1 f.seek(0)
         2
         3 for record in f:
         4     if '1' in record:
         5         print(record, end = '')
```

```
Line 1: 0.1 5.4 2 4 20 .03
Line 2: 1 3. 2.0 43 12
Line 3: 1 2
```

Read **one line** only: current line only

```
readline()
```

```
In [35]: 1 f.seek(0)
         2 record = f.readline()
```

```
In [36]: 1 print(record)
```

This line is 26 characters

```
In [37]: 1 s = f.readline()
```

```
In [38]: 1 print(s)
```

Line 1: 0.1 5.4 2 4 20 .03

Read ALL lines:

```
readlines()
```

```
In [39]: 1 f.seek(0)
         2 s = f.readlines()
         3 s
```

```
Out[39]: ['This line is 26 characters\n',
          'Line 1: 0.1 5.4 2 4 20 .03\n',
          'Line 2: 1 3. 2.0 43 12 \n',
          'Line 3: 1 2 \n',
          'This is the last record\n']
```

A list of strings!!!

Newline character is included!

```
In [40]: 1 for line in s:
          2     print(line, end = '')
```

```
This line is 26 characters
Line 1: 0.1 5.4 2 4 20 .03
Line 2: 1 3. 2.0 43 12
Line 3: 1 2
This is the last record
```

```
In [ ]: 1
```

2 Write data into a file

`.write(string)`

- To write something into a file, the open function must be invoked with one of the mode flags:
w, a, x, r+, w+, a+
- Opening with w erases file's content if file exists, writing starts at the (new) beginning
- Opening with a lets writing start at the end of the file (appending)
- Opening with r+ lets writing start at the beginning of the file (overwriting)
- If a file doesn't exist, w, a, x will create it
- The + versions allow both writing and reading

```
In [44]: 1 f = open('data2.txt', 'w')
          2 nbytes = f.write('New line: 0 3 5.5')
          3 nbytes = f.write('Another new line: 1 2 3')
          4
```

```
In [45]: 1 %cat data2.txt
```

```
New line: 0 3 5.5Another new line: 1 2 3
```

3 Close the file after use

Closing a file after use: `close()` method

➤ **Close** a file when no further operations are needed / allowed:

`file_handle.close()`

- Closing a file frees up used file resources (and let the file accessible for deleting/renaming by the OS)
- If a `close()` isn't explicitly called, python's garbage collector does eventually the job of closing the file
- Explicitly closing the file prevents the program to perform any (unwanted) further operations on file
- `close()` returns `None`

In [74]:

```
1 f = open('numbers.txt', 'w')
2 f.write('This file contains important data\n')
3 f.write('0 1 2 3 4\n')
4 f.write('4 3 2 1\n')
5 f.close()
```

In [75]:

```
1 s = f.read()
```

```
-----
-----
ValueError                                Traceback (most recent call
  last)
<ipython-input-75-5c4f2c28a2d3> in <module>
----> 1 s = f.read()

ValueError: I/O operation on closed file.
```

4 Dealing with errors

Dealing with errors: try-except-else-finally construct

- When an **error** occurs during the program, Python generates an **exception**: it generates an error type that identifies the exception and then **stops** the execution
- Exceptions can be handled using the **try statement** to avoid that the program does actually stop when an error occurs during the execution

- **try-except-else-finally blocks:**

- Optional
- ✓ The **try** block let executing a block of code that can potentially generate an exception
 - ✓ The **except** block let handling the error, if generated by the try block (i.e., what to do when an error occurs)
 - ✓ The **else** block let specifying a block of code that is executed if the try block *didn't generate any exception*
 - ✓ The **finally** block let executing the code, regardless of the result of the try- and except blocks.

```
y = 1
try:
    x /= 10
    y += x
except:
    print("x doesn't exist")
else:
    print('x:', x)
    del x
finally:
    print('y:', y)
```

Try / Except with files, simple version:

In [78]:

```
1 f = None
2 try:
3     filename = 'xyz.txt'
4     f = open(filename)
5 except:
6     print('File ' + filename + ' not found or not readable')
7 print(f)
```

File xyz.txt not found or not readable!
None

In [79]:

```
1 try:
2     filename = 'xyz.txt'
3     f = open(filename)
4 except:
5     print('File ' + filename + ' not found or not readable')
6     print('Opening numbers.txt ...')
7     f = open('numbers.txt')
8 print(f)
```

File xyz.txt not found or not readable!
Opening numbers.txt
<_io.TextIOWrapper name='numbers.txt' mode='r' encoding='UTF-8'>

5 Reading CSV files

CSV stands for **Comma Separated Values**. It's a quite flexible and compact for storing data.

It's around since long time, and it's the main format used by popular spreadsheet programs such as Excel.

Many data repositories make use of CSV as one their standard formats for data.

E.g., a CSV file with monthly evolution of the Amazon's stock market prices at Nasdaq from Yahoo! Finance:

<https://finance.yahoo.com/quote/AMZN/history?period1=1521362028&period2=1552898028&interval=1mo&filter=history&frequency=1mo>

Opening the file in **Excel**

[illegible]

Opening the file with a regular text editor/viewer

jupyter

AMZN.csv✓ a few seconds ago

FileEditViewLanguage

1

Date,Open,High,Low,Close,Adj Close,Volume

2

2018-04-01,1417.619995,1638.099976,1352.880005,1566.130005,1566.130005,129919600

3

2018-05-01,1563.219971,1635.000000,1546.020020,1629.619995,1629.619995,71553400

4

2018-06-01,1637.030029,1763.099976,1635.089966,1699.800049,1699.800049,85941300

5

2018-07-01,1682.699951,1880.050049,1678.060059,1777.439941,1777.439941,97521100

6

2018-08-01,1784.000000,2025.569946,1776.020020,2012.709961,2012.709961,96546400

7

2018-09-01,2026.500000,2050.500000,1865.000000,2003.000000,2003.000000,94445500

8

2018-10-01,2021.989990,2033.189941,1476.359985,1598.010010,1598.010010,183220800

9

2018-11-01,1623.530029,1784.000000,1420.000000,1690.170044,1690.170044,139290000

10

2018-12-01,1769.459961,1778.339966,1307.000000,1501.969971,1501.969971,154812700

11

2019-01-01,1465.199951,1736.410034,1460.930054,1718.729980,1718.729980,134001700

12

2019-02-01,1638.880005,1673.060059,1566.760010,1639.829956,1639.829956,80936900

13

2019-03-01,1655.130005,1709.430054,1651.000000,1668.949951,1668.949951,18811800

14

Example, a Kaggle dataset on car sales

< Car_sales.csv (15.64 KB)

DetailCompactColumn

16 of 16 columns

Sales announcements with multiple info

A Manufacturer	A Model	# Sales_in_thousands	# _year_resale_value	A Vehicle_Type	# Price_in_thousands	# Engine_size	# Horsepower	# Wheelbase	# Width	# Length	# Curb_weight
Dodge	Neon			Passenger							
Ford	Integra			Car							
Other (135)	Other (154)										
Acura	Integra	16.919	16.36	Passenger	21.5	1.8	148	181.2	67.3	172.4	2.639
Acura	TL	39.384	19.875	Passenger	28.4	3.2	225	188.1	70.3	192.9	3.517
Acura	CL	14.114	18.225	Passenger		3.2	225	186.9	70.6	192	3.47
Acura	RL	8.588	29.725	Passenger	42	3.5	210	114.6	71.4	196.6	3.85
Audi	A4	28.397	22.255	Passenger	23.99	1.8	158	182.6	68.2	178	2.998
Audi	A6	18.78	23.555	Passenger	33.95	2.8	280	188.7	76.1	192	3.561
Audi	A8	1.38	39	Passenger	62	4.2	310	113	74	198.2	3.982
BMW	323i	19.747		Passenger	26.99	2.5	170	187.3	68.4	176	3.179
BMW	328i	9.231	28.675	Passenger	33.4	2.8	193	187.3	68.5	176	3.197
BMW	528i	17.527	36.125	Passenger	38.9	2.8	193	111.4	70.9	188	3.472
Buick	Century	91.561	12.475	Passenger	21.975	3.1	175	189	72.7	194.6	3.368
Buick	Regal	39.35	13.74	Passenger	25.3	3.8	248	189	72.7	196.2	3.543
Buick	Park Avenue	27.851	28.19	Passenger	31.965	3.8	285	113.8	74.7	206.8	3.778
Buick	LeSabre	83.257	13.36	Passenger	27.885	3.8	285	112.2	73.5	208	3.591
Cadillac	DeVille	63.729	22.525	Passenger	39.895	4.6	275	115.3	74.5	207.2	3.978
Cadillac	Seville	15.943	27.1	Passenger	44.475	4.6	275	112.2	75	281	
Cadillac	Eldorado	6.536	25.725	Passenger	39.665	4.6	275	188	75.5	208.6	3.843
Cadillac	Catera	11.185	18.225	Passenger	31.81	3	280	187.4	70.3	194.8	3.77
Cadillac	Escalade	14.785		Car	46.225	5.7	255	117.5	77	281.2	5.572
Chevrolet	Cavalier	145.519	9.25	Passenger	13.26	2.2	115	184.1	67.9	188.9	2.676
Chevrolet	Malibu	135.126	11.225	Passenger	16.535	3.1	170	187	69.4	198.4	3.851
Chevrolet	Lumina	24.629	18.31	Passenger	18.89	3.1	175	187.5	72.5	208.9	3.33


```

1 Manufacturer,Model,Sales_in_thousands,__year_resale_value,Vehicle_type,Price_in_thousands,Engine_size,Horsepower,Whe
  elbase,Width,Length,Curb_weight,Fuel_capacity,Fuel_efficiency,Latest_Launch,Power_perf_factor
2 Acura,Integra,16.919,16.36,Passenger,21.5,1.8,140,101.2,67.3,172.4,2.639,13.2,28,2/2/2012,58.28014952
3 Acura,TL,39.384,19.875,Passenger,28.4,3.2,225,108.1,70.3,192.9,3.517,17.2,25,6/3/2011,91.37077766
4 Acura,CL,14.114,18.225,Passenger,,3.2,225,106.9,70.6,192,3.47,17.2,26,1/4/2012,
5 Acura,RL,8.588,29.725,Passenger,42,3.5,210,114.6,71.4,196.6,3.85,18,22,3/10/2011,91.38977933
6 Audi,A4,20.397,22.255,Passenger,23.99,1.8,150,102.6,68.2,178,2.998,16.4,27,10/8/2011,62.7776392
7 Audi,A6,18.78,23.555,Passenger,33.95,2.8,200,108.7,76.1,192,3.561,18.5,22,8/9/2011,84.56510502
8 Audi,A8,1.38,39,Passenger,62,4.2,310,113,74,198.2,3.902,23.7,21,2/27/2012,134.6568582
9 BMW,323i,19.747,,Passenger,26.99,2.5,170,107.3,68.4,176,3.179,16.6,26,6/28/2011,71.19120671
10 BMW,328i,9.231,28.675,Passenger,33.4,2.8,193,107.3,68.5,176,3.197,16.6,24,1/29/2012,81.87706856
11 BMW,528i,17.527,36.125,Passenger,38.9,2.8,193,111.4,70.9,188,3.472,18.5,25,4/4/2011,83.9987238
12 Buick,Century,91.561,12.475,Passenger,21.975,3.1,175,109,72.7,194.6,3.368,17.5,25,11/2/2011,71.18145132
13 Buick,Regal,39.35,13.74,Passenger,25.3,3.8,240,109,72.7,196.2,3.543,17.5,23,9/3/2011,95.63670253
14 Buick,Park Avenue,27.851,20.19,Passenger,31.965,3.8,205,113.8,74.7,206.8,3.778,18.5,24,3/23/2012,85.82840825
15 Buick,LeSabre,83.257,13.36,Passenger,27.885,3.8,205,112.2,73.5,200,3.591,17.5,25,7/23/2011,84.25452581
16 Cadillac,DeVille,63.729,22.525,Passenger,39.895,4.6,275,115.3,74.5,207.2,3.978,18.5,22,2/23/2012,113.8545976
17 Cadillac,Seville,15.943,27.1,Passenger,44.475,4.6,275,112.2,75,201,,18.5,22,4/29/2011,115.6213578
18 Cadillac,Eldorado,6.536,25.725,Passenger,39.665,4.6,275,108,75.5,200.6,3.843,19,22,11/27/2011,113.7658739
19 Cadillac,Catera,11.185,18.225,Passenger,31.01,3,200,107.4,70.3,194.8,3.77,18,22,9/28/2011,83.48309358
20 Cadillac,Escalade,14.785,,Car,46.225,5.7,255,117.5,77,201.2,5.572,30,15,4/17/2012,109.5091165
21 Chevrolet,Cavalier,145.519,9.25,Passenger,13.26,2.2,115,104.1,67.9,180.9,2.676,14.3,27,8/17/2011,46.36334747
22 Chevrolet,Malibu,135.126,11.225,Passenger,16.535,3.1,170,107,69.4,190.4,3.051,15,25,3/19/2012,67.31446216
23 Chevrolet,Lumina,24.629,10.31,Passenger,18.89,3.1,175,107.5,72.5,200.9,3.33,16.6,25,5/24/2011,69.9913956
24 Chevrolet,Monte Carlo,42.593,11.525,Passenger,19.39,3.4,180,110.5,72.7,197.9,3.34,17,27,12/22/2011,72.03091719
25 Chevrolet,Camaro,26.402,13.025,Passenger,24.34,3.8,200,101.1,74.1,193.2,3.5,16.8,25,10/23/2011,81.11854333
26 Chevrolet,Corvette,17.947,36.225,Passenger,45.705,5.7,345,104.5,73.6,179.7,3.21,19.1,22,5/12/2012,141.14115
27 Chevrolet,Prizm,32.299,9.125,Passenger,13.96,1.8,120,97.1,66.7,174.3,2.398,13.2,33,9/11/2011,48.2976361
28 Chevrolet,Metro,21.855,5.16,Passenger,9.235,1.55,93.1,62.6,149.4,1.895,10.3,45,4/13/2012,23.27627233
29 Chevrolet,Impala,107.995,,Passenger,18.89,3.4,180,110.5,73,200,3.389,17,27,6/18/2011,71.83803944
30 Chrysler,Sebring Coupe,7.854,12.36,Passenger,19.84,2.5,163,103.7,69.7,190.9,2.967,15.9,24,1/16/2012,65.95718396

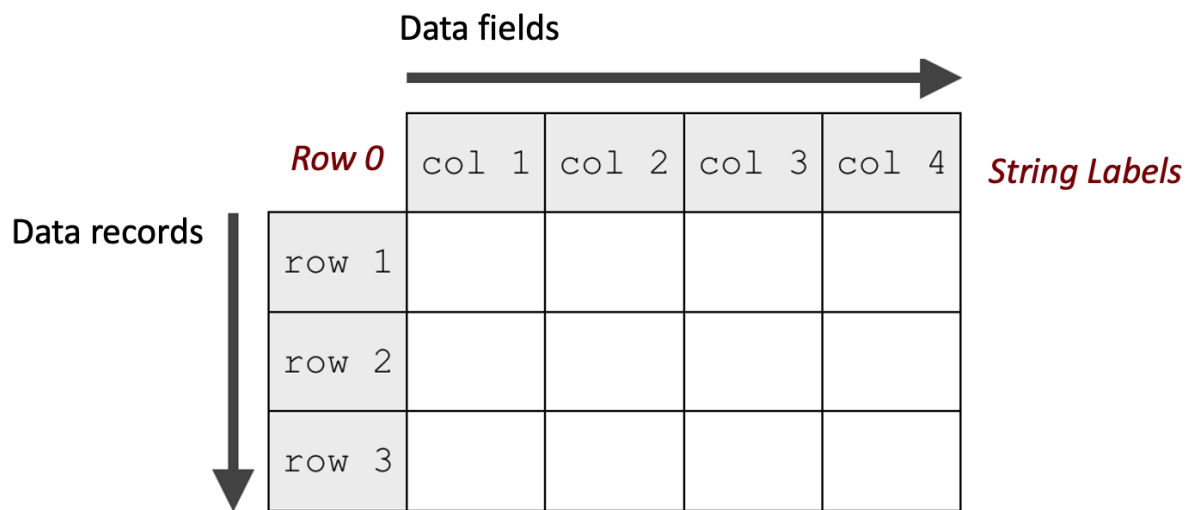
```

Datasets are commonly stored in **tabular form** (a *matrix* of data, a two-dimensional *data frame*): M data records/rows, each consisting of (at most) N ordered fields/columns

→ Each column/field describes the value of a property of interest.

→ Each row / record is a set of data making one **sample/example**.

→ The **first row**, optionally, can/should include **string labels**, explaining the meaning of the columns (i.e., giving a specific name to a column/attribute of data).



Both columns and rows (optionally) usually have explanatory **labels**, (e.g., 'Model', 'Manufacturer', 'Switzerland') or numeric indices (e.g., record #15) that are useful to refer to data, especially when manipulating them interactively.

Column data are separated by a given delimiter.

The default delimiter is a *comma*, but other characters can be used as a delimiter

→ The module `csv` makes available useful methods for dealing with CSV files

```
import csv
```

In [112]:

```
1 import csv
2
3 file_path = 'employee_addresses.csv'
4
5 f_csv = open(file_path)
6
7 csv_data = csv.reader(f_csv, delimiter=',')
8
```

```
In [101]: 1 csv_data
```

```
Out[101]: <_csv.reader at 0x7ff20a70fb38>
```

csv_data is an **iterator**: at each call will return the **next line** in the file

Data records are read into **lists of strings**, where each list element is a string with a value identified based on the given delimiter

```
In [109]: 1 line_count = 0
          2
          3 for row in csv_data:
          4     print(row)
          5     line_count += 1
```

```
['name', 'address', 'date joined']
['john smith', '1132 Anywhere Lane Hoboken NJ', ' 07030', 'Jan 4']
['erica meyers', '1234 Smith Lane Hoboken NJ', ' 07030', ' March 2']
['ann mcdonald', ' 9223 Yoda Lane Pythonopolis CA', ' 90001', 'April 1']
```

We can use `sep.join(row)` to get / print out the whole row as a string with the desired separator, `sep`

```
In [110]: 1 f_csv.seek(0)
          2
          3 line_count = 0
          4
          5 for row in csv_data:
          6     print(' '.join(row))
          7     line_count += 1
```

```
name address date joined
john smith 1132 Anywhere Lane Hoboken NJ 07030 Jan 4
erica meyers 1234 Smith Lane Hoboken NJ 07030 March 2
ann mcdonald 9223 Yoda Lane Pythonopolis CA 90001 April 1
```

Remember that the **first line**, usually represents a **header**, that specifies the names of the fields in the records.

A csv file can be seen as a **dictionary**: each column has a label, hence, we can read the csv data file (or, more generically, tabular data) into an ordered dictionary, an dictionary that preserves/remembers the order for entering the keys. The keys are sorted by the order associated to their entrance in the dictionary.

Each row is an ordered dictionary with respect to the keys/columns.

→ Need to get the iterator using the call to `csv.DictReader()`

In [113]:

```
1 file_path = 'employee_addresses.csv'
2
3 f_csv = open(file_path)
4
5 csv_data = csv.DictReader(f_csv)
```

With `csv_data.fieldnames` we get the content of the first row, that is, the field names into a list of strings.

In [114]:

```
1 print(csv_data.fieldnames)
['name', 'address', 'date joined']
```

How many fields / attributes / columns do we have in the dataset?

In [119]:

```
1 num_of_keys = len(csv_data.fieldnames)
2 print('The file has ' + str(num_of_keys) + ' fields')
```

The file has 3 fields

The really useful thing is that we can now retrieve the individual fields as in a dictionary, using the keys!

In [130]:

```
1 f_csv.seek(0)
2
3 r_num = 1
4 next(csv_data)  # this reads / advances one record, the
5 for r in csv_data:
6     print('Record n. ' + str(r_num) + ' - ' + 'Address i
7     r_num += 1
8
```

Record n. 1 - Address is: 1132 Anywhere Lane Hoboken NJ
Record n. 2 - Address is: 1234 Smith Lane Hoboken NJ
Record n. 3 - Address is: 9223 Yoda Lane Pythonopolis CA

Get the value of the name field if the address is in California

In [132]:

```
1 f_csv.seek(0)
2
3 r_num = 1
4 for r in csv_data:
5     if 'CA' in r['address']:
6         print('Record n. ' + str(r_num) + ' - ' + 'Name
7     r_num += 1
8
9
```

Record n. 4 - Name is: ann mcdonald

6 Read input from the keyboard

What about reading data directly from user input?

`input()`

In [135]:

```
1 n = input('Enter a number: ')
2
3 print('You have entered:', n)
```

Enter a number: 123
You have entered: 123

In [136]:

```
1 while True:
2     n = input('Enter a number: ')
3
4     if n.isnumeric():
5         print('You have entered:', n)
6         break
7     else:
8         print('Please enter a valid number!')
9
```

Enter a number: jadgf
Please enter a valid number!
Enter a number: 23
You have entered: 23

In [140]:

```
1 print('I\'ll make the sum of the number that you will enter!')
2 print('Enter 0 to end')
3
4 numbers = []
5 while True:
6     n = input('Enter a number: ')
7
8     if n == '0':
9         print('Sum is: ' + str(sum(numbers)))
10        break
11
12    if n.isnumeric():
13        numbers.append(int(n))
14    else:
15        print('Please enter a valid number!')
16
```

I'll make the sum of the number that you will enter!
Enter 0 to end
Enter a number: 2
Enter a number: 3
Enter a number: 5
Enter a number: 0
Sum is: 10

