Files Chris Piech and Mehran Sahami

诸子作

duss me

Who we are today



Galileo Galilei

Review



Dictionaries

Dictionaries allow us to build one-way associations between one kind of data (which we call the key) and another (which we call the value). A common metaphor for them is a phone book, whose keys are people's names and whose values are their phone numbers. It's super easy to look someone's number up if you know their name, but harder to do it the other way around.

What's a file?



Files on your computer



File Home Share	View		
n to Quick Copy Paste	Copy path Paste shortcut	ame New folder	access • Properties
Clipboard		New	Ope
• -> • 🛧 🛅 • TI	his PC > Documents >		
	Name	Date modified	Туре
Quick access Downloads ploads	atcp-profiles.zip	10-Aug-18 10:23 A	Compressed File (
	lee.jpg	10-Aug-18 09:53 A	IPG File
	apps.36009.13608414915328238.dfdb6209	10-Aug-18 09:43 A	IPG File
🗄 Documents 🚿	gphotos.jpg	10-Aug-18 09:20 A	JPG File
💪 Google Drive 🖈	photo-stacking.jpg	10-Aug-18 09:02 A	JPG File
	gmail-contacts-removal.png	10-Aug-18 08:32 A	PNG File
	inprivate-desktop.jpg	08-Aug-18 06:23 P	JPG File
This PC	New Database.odb	08-Aug-18 01:06 P	OpenDocument D
i Network	🔂 GotchaBackup.zip	07-Aug-18 01:59 P	Compressed File (
	😭 windows-10-hi-animation.zip	07-Aug-18 07:14 A	Compressed File (
	🐻 Thunderbird Setup 60.0.exe	06-Aug-18 12:28 P	Application
	WMPTagPlus-2.10.exe	06-Aug-18 09:34 A	Application





When you edit a word document...

... the program *Microsoft Word* is opening the file essay.docx



When you edit a word document...

... the program *Microsoft Word* is opening the file essay.docx

When you watch a video on your computer

... the program VLC is opening the file Avengers: Endgame.mp4



When you edit a word document...

... the program *Microsoft Word* is opening the file essay.docx

When you watch a video on your computer

... the program VLC is opening the file Avengers: Endgame.mp4

When you listen to a song...

... the program *iTunes* is opening the file Thriller.mp3



img = SimpleImage("buddy.jpg")

buddy.jpg





img = SimpleImage("buddy.jpg")





python3 program.py

program.py
def main():
 # Python code here
if ______ == "____main___":
 main()



python3 program.py

program.py

def main():
 # Python code here
if __name__ == "__main__":
 main()

The program python3 is opening the file program.py



How files are stored

foundation.txt

His name was Gaal Dornick and he was just a country boy who had never seen Trantor before...

program.py

def main():
 print("Hello world!")

thechain.mp3

spiderverse.mp4







How files are stored

foundation.txt

program.py

thechain.mp3

spiderverse.mp4





How are files stored?

foundation.txt

program.py

thechain.mp3

spiderverse.mp4

00101111 00111001 01101010 00101111 00110100 01010010 01001100 01101011 01010010 01011000...



How files are stored





How files are stored







foundation.txt

His name was Gaal Dornick and he was just a country boy who had never seen Trantor before...





foundation.txt

His name was Gaal Dornick and he was just a country boy who had never seen Trantor before...

Text files store data that you can read





foundation.txt	His name was Gaal Dornick and he was just a country boy who had never seen Trantor before	
countries.csv	Argentina,Buenos Aires,45M,South America Singapore,Singapore,5.7M,Asia New Zealand,Wellington,4.9M,Oceania	

Text files store data that you can read





foundation.txt	His name was Gaal Dornick and he was just a country boy who had never seen Trantor before	
countries.csv	Argentina,Buenos Aires,45M,South America Singapore,Singapore,5.7M,Asia New Zealand,Wellington,4.9M,Oceania	
locations.json	<pre>{"Chris": "Stanford", "Brahm": "New York"}</pre>	

Text files store data that you can read





foundation.txt	His name was Gaal Dornick and he was just a country boy who had never seen Trantor before	
countries.csv	Argentina,Buenos Aires,45M,South America Singapore,Singapore,5.7M,Asia New Zealand,Wellington,4.9M,Oceania	
locations.json	<pre>{"Chris": "Stanford", "Brahm": "New York"}</pre>	

Text files store data that a Python program can read







Where do values come from in Python?













Where do values come from in Python?





Where do values come from in Python?



Reading a file



invictus.txt

Out of the night that covers me, Black as the pit from pole to pole, I thank whatever gods may be For my unconquerable soul. . . . I am the master of my fate, I am the captain of my soul.

'Invictus', by William Ernest Henley





```
print_poem.py
```

```
f = open('invictus.txt')
for line in f:
    # process the line
f.close()
```

invictus.txt

Out of the night that covers me, Black as the pit from pole to pole, I thank whatever gods may be For my unconquerable soul. . .

I am the master of my fate, I am the captain of my soul.

'Invictus', by William Ernest Henley







print poem.py
















3.5 stanzas later...















f

line

Reading a file in Python



→ "'Invictus', by William Ernest Henley"

X

invictus.txt

Out of the night that covers me, Black as the pit from pole to pole, I thank whatever gods may be For my unconquerable soul. .

I am the master of my fate, I am the captain of my soul.

'Invictus', by William Ernest Henley

🖬 Files 🛛 🕂	
🗎 invictus.txt	



f

line

Reading a file in Python



¯_(ツ)_/¯

-

"'Invictus', by William Ernest Henley"

invictus.txt

Out of the night that covers me, Black as the pit from pole to pole, I thank whatever gods may be For my unconquerable soul. . . . I am the master of my fate,

I am the captain of my soul.

'Invictus', by William Ernest Henley

Files +
🗎 invictus.txt







Reading a file in Python: Newlines



Computers don't know that 'lines' are shown below one another in a file.

ut of the night that covers me,
Black as the Pit from pole to pole,
thank whatever gods may be
For my unconquerable soul.
n the fell clutch of circumstance
I have not winced nor cried aloud.
nder the bludgeonings of chance
My head is bloody, but unbowed.
eyond this place of wrath and tears
Looms but the Horror of the shade,
nd yet the menace of the years
Finds, and shall find, me unafraid.
t matters not how strait the gate,
How charged with punishments the scroll,
am the master of my fate:
I am the captain of my soul.

'Invictus', by William Ernest Henley



They store all the lines in a file in one continuous block...

Out of the night that covers me, Black as the Pit from pole to pole, I thank whatever gods may be For my unconquerable soul.In the fell clutch of circumstance I have not winced nor cried aloud. Under the bludgeonings of chance My head is bloody, but unbowed.Beyond this place of wrath and tears Looms but the Horror of the shade, And yet the menace of the years Finds, and shall find, me unafraid.It matters not how strait the gate, How charged with punishments the scroll, I am the master of my fate: I am the captain of my soul.'Invictus', by William Ernest Henley



Reading a file in Python: Newlines



...and separate the lines with a special newline character, n.

Out of the night that covers me, \n Black as the Pit from pole to pole, \nI thank whatever gods may be\n For my unconquerable soul. \nIn the fell clutch of circumstance\n I have not winced nor cried aloud.\nUnder the bludgeonings of chance\n My head is bloody, but unbowed. \nBeyond this place of wrath and tears \n Looms but the Horror of the shade, \nAnd yet the menace of the years \n Finds, and shall find, me unafraid. \nIt matters not how strait the gate, \n How charged with punishments the scroll, \nI am the master of my fate: \n I am the captain of my soul.\n\n'Invictus', by William Ernest Henley



```
print_poem.py
```

```
f = open('invictus.txt')
for line in f:
    print(line)
f.close()
```

invictus.txt

'Invictus', by William Ernest Henley













print(line)

line \longrightarrow "Out of the night that covers me,\n"



print("Out of the night that covers me,\n")



print("Out of the night that covers me,\n") print(" Black as the Pit from pole to pole,\n")







This \n comes from the string we're printing

This \n comes from the print function

\n



print("Out of the night that covers me,\n") print(" Black as the Pit from pole to pole,\n")

Out of the night that covers me,\n \n Black as the Pit from pole to pole,\n \n



A file reading loop

```
f = open(filename)
for line in f:
    line = line.strip() # or strip("\n"), lstrip(), rstrip()
    # process line
f.close()
```



```
with open(filename) as f:
    for line in f:
        line = line.strip() # or strip("\n"), lstrip(), rstrip()
        # process line
```



The file reading loop





Skipping lines in a file

```
invictus.txt
print_poem.py
                                                   Out of the night that covers me,
 f = open('invictus.txt')
                                                          Black as the pit from pole to pole,
 next(f)
                                                    I thank whatever gods may be
 for line in f:
                                                          For my unconquerable soul.
      # process the line
 f.close()
                                                    I am the master of my fate,
                                                          I am the captain of my soul.
                                                     'Invictus', by William Ernest Henley
  f
```

invictus.txt



Skipping lines in a file



invictus.txt



Skipping lines in a file













f.read() and f.readlines() are handy, but shouldn't be your first resort. If you try to read really big files, that takes up a lot of your computer's memory, which you should try to avoid. You generally won't need the contents of the entire file as a string or list, anyway.

Writing a file






























Writing to files





Writing to files



Visualizing Data



What can we do with data in a file?





What can we do with data in a file?







An interesting data file

United States.csv

City,Latitude,Longitude,Population New York,40.6943,-73.9249,18713220 Los Angeles,34.1139,-118.4068,12750807 Chicago,41.8373,-87.6862,8604203 Miami,25.7839,-80.2102,6445545 Dallas,32.7936,-96.7662,5743938 Philadelphia,40.0077,-75.1339,5649300 Houston,29.7863,-95.3889,5464251

- •
- •
- .



Defining some terms



-90



An interesting data file

United States.csv

City,Latitude,Longitude,Population New York,40.6943,-73.9249,18713220 Los Angeles,34.1139,-118.4068,12750807 Chicago,41.8373,-87.6862,8604203 Miami,25.7839,-80.2102,6445545 Dallas,32.7936,-96.7662,5743938 Philadelphia,40.0077,-75.1339,5649300 Houston,29.7863,-95.3889,5464251





An interesting data file

United States.csv

City,Latitude,Longitude,Population New York,40.6943,-73.9249,18713220 Los Angeles,34.1139,-118.4068,12750807 Chicago,41.8373,-87.6862,8604203 Miami,25.7839,-80.2102,6445545 Dallas,32.7936,-96.7662,5743938 Philadelphia,40.0077,-75.1339,5649300 Houston,29.7863,-95.3889,5464251 Key idea: Represent each city as a pixel on an image





foundation.txt	His name was Gaal Dornick and he was just a country boy who had never seen Trantor before
countries.csv	Argentina,Buenos Aires,45M,South America Singapore,Singapore,5.7M,Asia New Zealand,Wellington,4.9M,Oceania
locations.json	{"Chris": "Stanford", "Brahm": "New York"}

.csv and .json files are more structured than arbitrary text files



foundation.txt	His name was Gaal Dornick and he was just a country boy who had never seen Trantor before
countries.csv	Argentina, Buenos Aires, 45M, South America Singapore, Singapore, 5.7M, Asia New Zealand, Wellington, 4.9M, Oceania
locations.json	<pre>{"Chris": "Stanford", "Brahm": "New York"}</pre>

We *could* manually open a **csv** or **json** file and read and process it, but that would be annoying and lead to a lot of repeated code...



foundation.txt	His name was Gaal Dornick and he was just a country boy who had never seen Trantor before
countries.csv	Argentina,Buenos Aires,45M,South America Singapore,Singapore,5.7M,Asia New Zealand,Wellington,4.9M,Oceania
locations.json	<pre>{"Chris": "Stanford", "Brahm": "New York"}</pre>

...so Python gives us utilities for working with them!



Reading csv files

USA.csv

City,Latitude,Longitude,Population New York,40.6943,-73.9249,18713220 Los Angeles,34.1139,-118.4068,12750807 Chicago,41.8373,-87.6862,8604203 Miami,25.7839,-80.2102,6445545 Dallas,32.7936,-96.7662,5743938 Philadelphia,40.0077,-75.1339,5649300 Houston,29.7863,-95.3889,5464251

plot countries.py import csv def plot country(): with open("USA.csv") as f: next(f) reader = csv.reader(f) for line in reader: lat = float(line[1]) lon = float(line[2])

Looping through a csv.reader splits by commas for you



Reading csv files

USA.csv

```
City,Latitude,Longitude,Population
New York,40.6943,-73.9249,18713220
Los Angeles,34.1139,-118.4068,12750807
Chicago,41.8373,-87.6862,8604203
Miami,25.7839,-80.2102,6445545
Dallas,32.7936,-96.7662,5743938
Philadelphia,40.0077,-75.1339,5649300
Houston,29.7863,-95.3889,5464251
```

```
plot_countries.py
```

import csv

```
def plot_country():
    with open("USA.csv") as f:
        reader = csv.DictReader(f)
```

```
for line in reader:
    lat = float(line['Latitude'])
    lon = float(line['Longitude'])
```

Looping through a csv.DictReader and gives you back each line as a dictionary whose keys are column names and whose values are that column's value for that particular row



Writing csv files



A csv.writer allows you to write rows as lists of data, without worrying about things like \n characters



Writing csv files

data.csv	write_data.py
x,y 1,2 2,4 4,6	<pre>import csv def write_data(): with open("data.csv", "w") as f: columns = ['x', 'y'] writer = csv.DictWriter(f) writer.writeheader(fieldnames=columns) writer.writerow({'x': 1, 'y': 2}) writer.writerow({'x': 2, 'y': 4}) writer.writerow({'x': 4, 'y': 6})</pre>

A csv.DictWriter allows you to write rows as dictionaries of data, without worrying about things like \n characters



Why use csv files?

- Great for tabular data
- Understood by Excel and Google Sheets!



Reading json files

ages.json

```
{
    "Mehran": 18,
    "Julie": 18,
    "Brahm": 24,
    "Chris": 33
}
```

\$ python3 read_ages.py
Brahm is 24

```
read_ages.py
import json
def get_ages():
    with open("ages.json") as f:
        ages = json.load(f)
        age = ages['Brahm']
        print("Brahm is " + str(age))
```

json.load allows you to read a dictionary from a json file without having to manually process the file



Reading json files

to_buy.json

```
["Sugar", "Flour", "Salt"]
```

```
$ python3 shopping_list.py
["Sugar", "Flour", "Salt"]
3
```

shopping_list.py

import json

```
def get_shopping_list():
    with open("to_buy.json") as f:
        shopping_list = json.load(f)
        print(shopping_list)
        print(len(shopping_list))
```

json.load allows you to read a list from a json file without having to manually process the file



Writing json files



json.dump allows you to easily write a dictionary to a json file. The indent=4 parameter makes the file more readable)



Writing json files

data.json

```
["Rorschach", "Silk Spectre"]
```

```
shopping_list.py
import json
def write_data():
    data = ["Rorschach", "Silk Spectre"]
    with open("data.json", "w") as f:
        json.dump(data, f, indent=4)
```

json.dump allows you to easily write a list to a json file.



Why use json files?

- Great for structured data that *isn't* a table
- Will convert numbers to strings and back again
- It's the language of the internet
 - JSON stands for JavaScript Object Notation, and JavaScript is one of the languages that the internet uses
- Come back next time for more...