SI 506: Programming I Fall 2019

Lecture 18

Anthony Whyte <arwhyte@umich.edu>
Lecturer III, School of Information
715 N. University Ave, Ann Arbor, MI 48109
Roumanis Square, 2nd floor ("the loft")





preliminaries





Exercises paths

```
lecture_18_exercise.py
lecture_18_exercise_solution.py
un-americas_country_areas_dev.csv
un-americas_country_areas.csv
un-americas_regions_country_areas.csv
world_bank_americas_economies.csv
```





Data sets

UN regional, country/area data, and World Bank data

Countries/areas comprising the Americas

- Northern America
- Caribbean
- Central America
- South America







Contract Deliverables

- Process four *.csv files of UN/World Bank country/area data. Retain special characters.
- 2. Return counts of country/areas by region (data structure specified).
- 3. Write counts to *.json file.
- 4. Combine individual country/area data (data structure specified, exclude null values).
- 5. Write combined country/area data to *.json file.





region counts

```
"regions": {
  "americas": 57
"subregions": {
  "latin_america_and_the_caribbean": 52,
  "northern america": 5
"intermediate_regions": {
  "caribbean": 28,
  "central_america": 8,
  "south america": 16
```





new record (semi-structured)

```
"country_area": "Brazil",
"iso_alpha3_code": "BRA",
"locations": {
  "region": "Americas",
  "subregion": "Latin America and the Caribbean",
  "intermediate_region": "South America"
},
"development_status": {
  "un_classification": "Developing",
  "world_bank_classification": "Upper middle income economy"
}
```





new record (semi-structured)

```
"country_area": "Canada",
"iso_alpha3_code": "CAN",
"locations": {
  "region": "Americas",
                                           no intermediate region
  "subregion": "Northern America"
},
"development_status": {
  "un_classification": "Developed",
  "world_bank_classification": "High income economy"
}
```





new record (semi-structured)

```
"country_area": "Saint Barthélemy",
                                          retain special characters
"iso_alpha3_code": "BLM",
"locations": {
  "region": "Americas",
  "subregion": "Latin America and the Caribbean",
  "intermediate_region": "Caribbean"
},
"development_status": {
  "un_classification": "Developing"
                              no World Bank economic classification
```





Data set I UN 'Americas' countries/areas

file name: un-americas_country_areas.csv

file type: comma-delimited text file (.csv)

delimiter: comma (',')

row count: 58

header row: yes

row values: country/area, ISO Alpha3 code

character set: UTF-8

special characters: yes

description: countries/areas comprising the Americas.





Data set II

UN 'Americas' countries/territories development status

file name: un-americas_country_areas_dev_status.csv

file type: comma-delimited text file (.csv)

delimiter: comma (',')

row count: 58

header row: yes

row values: country/area, development status

character set: UTF-8

special characters: yes

description: countries/areas comprising the Americas classified as

either developing or developed.





Data set III

UN 'Americas' regions and countries/areas

file name: un-americas_regions_country_areas.csv

file type: comma-delimited text file (.csv)

delimiter: comma (',')

row count: 58

header row: yes

row values: region, subregion, intermediate region, country/area

character set: UTF-8

special characters: yes

description: regional affiliations for countries/areas comprising the Americas. Not all country/areas assigned to an intermediate region.





Data set IV

World Bank country/area economy classification

file name: world_bank-americas_economies.csv

file type: comma-delimited text file (.csv)

delimiter: comma (',')

row count: 58

header row: yes

row values: ISO Alpha 3 code, economic classification

character set: UTF-8

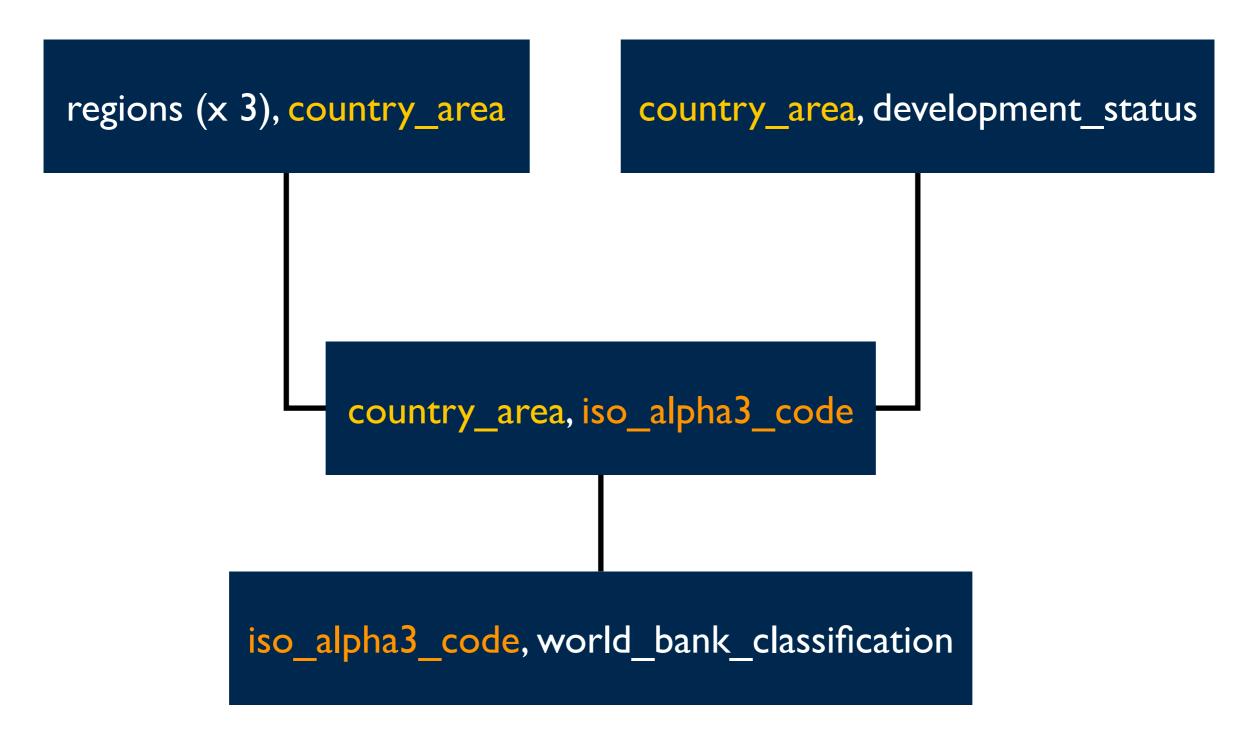
special characters: yes

description: national economies in the Americas classified as high income, upper middle income, lower middle income, or low income.





Data connections







lecture_18_exercise.py

script anatomy

```
load
```

```
import csv
import json
```

```
CONSTANTS
```

entry

call

```
main() process, delegate tasks
```

```
create_record(<5 parameters>)
format_keys(string)
get_region_count(data, region)
get_country_area_regions(regions, headers, country_area)
get_dev_status(dev_status, country_area)
get_economic_status(ec_status, iso_code)
read_csv(filename, seq_type=SEQ_TYPE[0])
write_json(filename, data)
```

```
if __name__ == '__main__':
    main() # call main method
```





CONSTANTS

putting a tuple to good use

```
ca_count = get_region_count(regions, INTERMED_REGIONS[1])
```





Why create a function?

enhance modularity, reusability, avoid duplicate code

```
for country_area in country_areas:
    for region in regions:
        if region[-1] == country_area_name:
            locations = {region_headers[0]: region[0],
                         region_headers[1]: region[1]}
            if region[2]:
                locations[region_headers[2]] = region[2]
            new_record['locations'] = locations
            break ←
                                               Yuck!
```





Function: get regional affiliations

returns regional affiliations

no break statement required





exercise





finis





directors cut





VS Code

opening files from a script

I. use absolute paths

```
import os
this_dir = os.path.dirname(os.path.abspath(__file__))
os.path figures out the current folder that points to the file you're running for any OS you're on.
file_to_open = os.path.join(this_dir, 'somefile.csv')
```

- 2. change the python.terminal.executeInFileDir
 - I. Search for "python.terminal.executeInFileDir"
 - 2. Check the check box (change vales from False to True
 - 3. Close settings window





Tuples data type characteristics

- sequence: siblings: list, string
- ordered: access items via indexing, slicing
- immutable: item reassignment not permitted
- comparable: comparison operator friendly
- hashable: can be used as dictionary keys





When your code misbehaves debug flowchart

Attribute Error

You are calling a method on the wrong type of object

SyntaxError

You've forgotten the quotes around a string

You have forgotten to put a colon at the end of a def/if/for line

You have different number of open and close brackets in a statement

TypeError

You're trying to use an operator on the wrong type of objects

An object which you expect to have a value is actually None

You've used non-integer numbers in a list slice

You've called a method/ function with the wrong number or type of arguments

Indentation Error

You've used a mixture of tabs and spaces You haven't indented all

lines in a block equally

My code isn't working:-(

Start here...

Do you get an

error when you

run the code?

Does the code

use loops or if

statements?

Two numbers which should

be equal are not

You are comparing a number

with a string representation

of a number (e.g. if 3 == "3")

A complex condition is not

giving the expected result

The order of precedence in the

condition is ambiguous - add

some parentheses

What type of error do you get?

NameError

You've misspelt a variable, function or method name

> You've forgotten to import a module

> You've forgotten to define a variable

Your code uses a variable outside the scope where it's defined

Your code calls a function before it's defined

You're trying to print a single word and have forgotten the quotes

IOError

You're trying to open a file that doesn't exist

KeyError

You're trying to look up a key that doesn't exist in a dict

http://pythonforbiologists.com

A variable that should contain a value does not

You are storing the return value of a function which You are printing an object changes the variable itself (e.g. sort)

A number which should be a fraction is coming out as zero in Python 2

You are dividing integers rather than floats. Convert the numbers to floats or from __future__ import division

I'm trying to print a value but getting a weirdlooking string

(e.g. a FileObject) when you want the result of calling a method on the object

A regular expression is not matching when I expect it to

You have forgotten to use raw strings or escape backslash characters

I am reading a file but getting no input

You have already read the contents of the file earlier in the code, so the cursor is at the end.

neithei

loops

A list which should have a value for every iteration only has a single value

You have defined the list inside the loop: move it outside

A loop which uses the range function misses out the last value

The range function is exclusive at the finish: increase it by one.

I am trying to loop over a collection of strings, but am getting individual characters

You are iterating over a string by mistake

I am trying to write multiple lines to a file but only getting a single one You have opened the file inside the loop: move it

also check.





Slide deck revisions

errata: corrections and other changes

Slide no(s). Fix ver. Description
v1p1



