HW5: CS 110X C 2014

Note: This homework (and all remaining homework assignments) is a **partner homework** and must be completed by each partner pair. When you complete this assignment, you must not share your answers with any other student. Only one person from a partner pair needs to submit the assignment, but make sure that you submit before the deadline!

For this assignment, every function that you write must have a suitable documentation string as we present in class. Check the rubric to see the point values assigned for each question so you can maximize the points you get on this assignment.

Please make sure that when you submit your assignment, you submit a single "HW5.py" file that contains your entire assignment.

Domain Information

You will find it useful to use data to support your decision making processes. For this homework assignment, you are to work with Comma Separated Value (CSV) files. These are the common format by which information is shared in a number of disciplines (science, business, engineering to name a few). A comma separated file contains a number of lines of text. The first line contains the **header** values that describe the names of the **fields** that will be represented as **columns** in the file. Each **row** is an entry in the file which contains specific values for each column. For example:

Quarter	Stores	Stores Closed	New Stores
2014Q1	992	20	14
2013Q4	998	6	16
2013Q3	988	4	18
2013Q2	970	0	16
2013Q1	954	8	12
2012Q3	950	11	8

```
Quarter,Stores,Stores Closed,New Stores
2014Q1,992,20,14
2013Q4,998,6,16
2013Q3,988,4,18
2013Q2,970,0,16
2013Q1,954,8,12
2012Q3,950,11,8
```

The above spreadsheet view (familiar to Excel users) would be represented by the CSV file shown above.

Canopy Issues

If you are running Canopy then you have to make a small configuration change for this homework to work properly. From within the Canopy Editor, select menu item **Edit | Preferences…**.
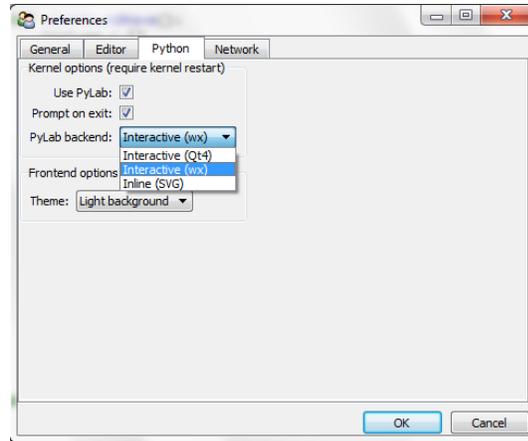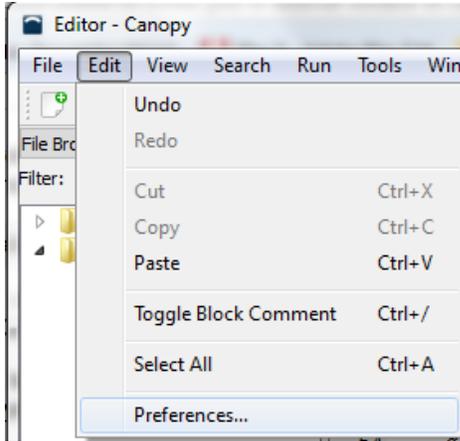
Then in the Preferences window, select the **Python** tab and be sure that **PyLab backend** is set to "Interactive (wx)". I have tested this on Windows and on a MacBook.

In Canopy, you need to set the working directory within the code editor. Near the right edge about one inch from the bottom you will see a small triangle that you can click on to change the working directory. Select the "Change to Editor Directory" so you will find the files that you need for this assignment.
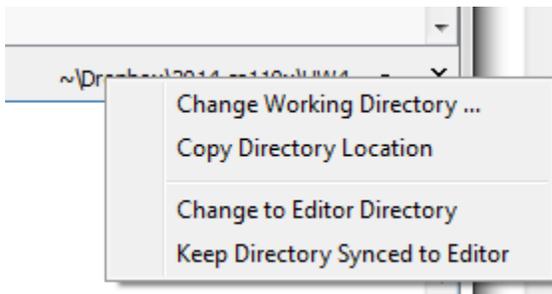
Homework Themes

This homework will exercise your skills in **CSV files** and the **dictionary** types.

Homework Instructions

This Homework has **Six** questions.

For each question be sure you understand exactly the format of the output that is requested. You will lose points if you do not exactly follow the format of the output for the individual questions. Should you have any questions, be sure to review the HW4 rubric and post questions on the HW4 discussion forum.

Q1	Demonstrate ability to work with lists
Skills CS-1 CS-3 PF-3	Write a function `fold(list1, list2)` that **returns** a single list containing paired elements, one from `list1` and one from `list2`. You can assume that `list1` and `list2` have the same length. Write a function `unfold(values)` that takes a list `values` of paired elements and **returns** a list `[list1, list2]` where `list1` is a list containing the first element in each pair while `list2` is a list containing the second element in each pair.
| Sample Output in IDLE | ```
>>> fold(['A', 'B', 'C'], [1, 2, 3])
[['A', 1], ['B', 2], ['C', 3]]
>>> unfold([['A', 1], ['B', 2], ['C', 3]])
[['A', 'B', 'C'], [1, 2, 3]]
``` |

Review the lecture on February 13th to become familiar with the Sacramento CSV data sets. In particular, use the `extractAllRecords(filename)` function to load up a list of rows from the CSV file. This function is already included in the HW5.py template file.

- SacramentocrimeJanuary2006.csv
- Sacramentorealestatetransactions.csv

Q2	Demonstrate knowledge of for loop, csv files, and list functions
Skills TBA *Lecture Dependency* *Feb-13*	Write a `filterRecords(listOfRows, fieldName, value)` function that processes the list of rows from a CSV file (header information and all) and **returns** a new list of rows that contains the header row and only those rows for which the field identified by `fieldName` **contains** the given string literal `value`. Note that this function **must always return at least one row**, namely, the header row. Also observe that you can use the **in** operator to determine whether a value is a substring of another string (as shown below in the sample output)

Sample Input File (smallDataSet.csv)

`Quarter,Stores,Stores Closed,New Stores`
`2014Q1,992,20,14`
`2013Q4,998,6,16`
`2013Q3,988,4,18`
`2013Q2,970,0,16`
`2013Q1,954,8,12`
`2012Q3,950,11,8`

| **Sample Output** | `>>> '2013' in '2013Q4'`
`True`
`>>> '2013' in '2014Q1'`
`False`
`>>> filterRecords(extractAllRecords('smallDataSet.csv'), 'Quarter', '2013')`
`[['Quarter', 'Stores', 'Stores Closed', 'New Stores'],`
` ['2013Q4', '998', '6', '16'],`
` ['2013Q3', '988', '4', '18'],`
` ['2013Q2', '970', '0', '16'],`
` ['2013Q1', '954', '8', '12']]`
`>>> filterRecords(extractAllRecords('smallDataSet.csv'), 'Stores', '1988')`
`[['Quarter', 'Stores', 'Stores Closed', 'New Stores']]` |

Q3	Demonstrate knowledge of for loop
Skills TBA **Lecture Dependency** Jan-30	Define a `positions`(`values, targets`) function that takes a `values` list and a `targets` list. It **returns** a list of integers corresponding to index locations within `values` for every target in `targets`. If any target in `targets` does not exist within `values`, then the corresponding index returned shall be `None`.
Sample Output in IDLE	```>>> positions(['a', 'b', 'c'], ['c', 'a'])``` ```[2, 0]```
Sample Output in Canopy	```In[1]: positions(['a', 'b', 'c'], ['d', 'a'])``` ```Out[1]: [None, 0]```

Q4	Demonstrate knowledge of dictionary data structure

Skills
 TBA
 	
 Lecture Dependency
 Feb-13
	Recall that in homework 3 that you used a function to determine the day of the week for a given year (1970 to 2099).
	
	 ```def januaryFirst(year):```
	 ``` year = year – 1```
	 ``` return (year + year/4 – year/100 + year/400 + 1) % 7```
	
	 Write a method `daysOfWeek(low,high)` that counts each of the days of the week of January 1st for each year in the period from `low` **up to and including** `high`. This function shall **return** a list containing seven integers representing the totals. The first value in the list is the total number of years in the designated period that started on Sunday.
	
	 Use a dictionary to solve this problem.

Sample Output in IDLE	```>>> daysOfWeek(2014, 2014):```
	 ```[0, 0, 0, 1, 0, 0, 0]```
	 ```>>> daysOfWeek(2000, 4000):```
	 ```[290, 280, 290, 285, 285, 290, 281]```

Q5	Demonstrate knowledge of for loop, csv, list functions
Skills TBA *Lecture Dependency* *Feb-13* *Feb-14*	For this question you are to write an `extractFields(listOfRows, listOfFields)` function that processes the list of rows from a CSV file (header information and all) and returns a new list that contains the same number of rows as the input `listOfRows` but only contains the header and column values for the fields identified in the `listOfFields` parameter. Note that this function **must always return** the same number of rows as `listOfRows`.

Sample Input File (smallDataSet.csv)

`Quarter,Stores,Stores Closed,New Stores`
`2014Q1,992,20,14`
`2013Q4,998,6,16`
`2013Q3,988,4,18`
`2013Q2,970,0,16`
`2013Q1,954,8,12`
`2012Q3,950,11,8`

| **Sample Output** | `>>> extractFields(extractAllRecords('smallDataSet.csv'), ['Stores'])`
`[['Stores'], ['992'], ['998'], ['988'], ['970'], ['954'], ['950']]`
`>>> extractFields(extractAllRecords('smallDataSet.csv'), ['Quarter', 'Stores'])`
`[['Quarter', 'Stores'],`
` ['2014Q1', '992'],`
` ['2013Q4', '998'],`
` ['2013Q3', '988'],`
` ['2013Q2', '970'],`
` ['2013Q1', '954'],`
` ['2012Q3', '950']]` |

Q6	Demonstrate knowledge of for loop, csv, list functions
Skills TBA *Lecture Dependency* *Feb-13* *Feb-14*	The final question for this assignment pulls together all parts to support the analyst working with the real estate and crimes data sets. Write a function `report(longitude, latitude, crimeRows, distance)` that allows an analyst to summarize the crimes that occur within a given distance of a specific (longitude, latitude) coordinate. Use a dictionary object to group these events by ' crimedescr' which is the field in the crimeRows data set. Use the `distanceMiles(lat1, long1, lat2, long2)` function provided in the template. This is a capstone question for this assignment

Sample Investigations that you can run with this function written	
Home in downtown Sacramento Note that all of these crimes are within ¼ mile of the given long/lat position	>>> report (-121.434879, 38.631913, extractAllRecords(crimesFileName), 0.25) 4 459 PC BURGLARY RESIDENCE 1 BATTERY - I RPT 1 MISCELLANEOUS I RPT (ZMISC) 1 TRESPASS OR PROWLER- I RPT 2 NARCOTICS SUSP/EVID/ACT- I RPT 2 243(E)1 BATTERY NONCOHAB SPOUS 5 FOUND PROPERTY - I RPT 1 11357(B)HS POSS -28.5GR MARIJ 1 487(A) GRAND THEFT-INSIDE 1 SUSP PERS-NO CRIME - I RPT 6 TOWED/STORED VEH-14602.6 2 10851(A)VC TAKE VEH W/O OWNER 1 O/S AGENCY -ASSISTANCE- I RPT 1 23152(A) VC DUI-ALCOHOL/DRUGS 1 TRAFFIC-ACCIDENT INJURY 1 12025(A)1)CONCLD GUN VEH/FEL 2 TOWED/STORED VEHICLE 1 3056 PAROLE VIO - I RPT 1 484 PETTY THEFT/LICENSE PLATE 1 TRAFFIC - I RPT 1 484G(B) PC ACCESS CARD FRAUD 2 MISSING PERSON 1 451(D) PC ARSON OF PROPERTY
Home in Wealthy Suburbs Note that you have to go 11 miles to find a close crime in the data set	>>> report (-121.219142, 38.732096, extractAllRecords(crimesFileName), 11) 1 VANDALISM - I RPT 1 5150 WI DANGER SELF/OTHERS 1 TRAFFIC - I RPT 1 459 PC BURGLARY VEHICLE 1 594(B)(1)PC VANDALISM +$400 1 594(B)(2)(A) VANDALISM/ -$400 1 242 PC BATTERY CIVILIAN 43 MISSING PERSON 1 1203.2 PC VIOLATION OF PROBATI 1 CHILD WELFARE - I RPT

Note: These outputs have been corrected from the previous version assignment…

How To Get Started On This Assignment

A template HW5.py file is provided to you with some sample functions already provided.

Submit your HW5.py file using the web-based turnin system. As we have mentioned in class, only one of the team members needs to submit the assignment. But just make sure that something gets submitted!

Change Log

1. For Q3 make sure you write a function **positions** not the name "position"
2. For Q6, note that I had my longitude/latitude coordinates reversed in the example, so I have generated the proper outputs
3. The CSV data sets were generated on a PC and therefore cause problems with Macintosh devices. I have fixed these and re-uploaded CSV files that work on both platforms. Please download the files again to make sure you are using the proper data sets.
4. For question 6, nothing is returned. Rather the output is printed to the screen
5. For question 6, note that the report method takes four parameters, with the last one being a distance value in miles.
6.