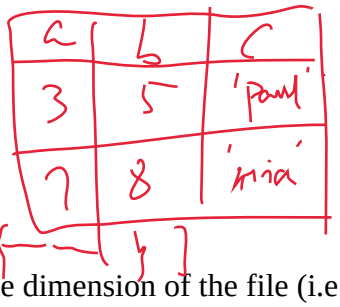
More CSV File Processing



CSV file

- Use the get_csv function to obtain a list of dictionaries.
- Make sure that you use the hierarchical breakdown to choose index or key
- (nested) for loop is useful to go through the data

d = [{'a': 3, 'b': 5, 'c': 'paul'}, {...}]

Coding Challenge

1. Write a function that takes in a list of dictionaries from a csv file, and returns the dimension of the file (i.e. number of rows and number of columns). The dimension should be returned as a tuple of 2 elements

```
def get_dim(data):
```

for x in d:
x is a dict

```
from CSE8ACSV import *
def get_dim(data):
    num_rows = len(data)
    num_cols = len(data[0])
    return (num_rows, num_cols)

d = get_csv('tech_diversity.csv')
print(get_dim(d))
```

2. Write a function that takes in the data from tech_diversity.csv and calculate the average of people_of_color percentage for all the companies.

```
def people_of_color_ave(data):
```

var for total
# of rows
for each row:
get the value for 'people_of_color' (type cast to float)
add it to total
divide total by # of rows
return the average

```
from CSE8ACSV import *
def people_of_color(data):
    num_rows = len(data)
    total = 0
    for row in data:
        total += float(row['people_of_color'])
    return total/num_rows

d = get_csv('tech_diversity.csv')
print(people_of_color(d))
```

3. Write a function that takes in the data from tech_diversity.csv and return the best company for each category. The return value should be a dictionary with the format of category:company name.

```
def find_best(data):
```

{'hispanic': 'Apple', 'white': 'Google', ... }

start w/ an empty dict
foreach key in the file: < How to handle 'company'
    best Val = 0
    declare a company name variable
    for each row in data:
        get the value for the key (type cast)
        if val is better than best Val:
            best Val is changed to val
            remember the company name
    insert key company into dict
return dict

```
from CSE8ACSV import *
def find_best(data):
    result = {}
    for key in data[0]:
        if key == 'Company':
            continue
        bestVal = 0
        company = ''
        for row in data:
            val = float(row[key])
            if val > bestVal:
                bestVal = val
                company = row['Company']
        result[key] = company
    return result

d = get_csv('tech_diversity.csv')
print(find_best(d))
```