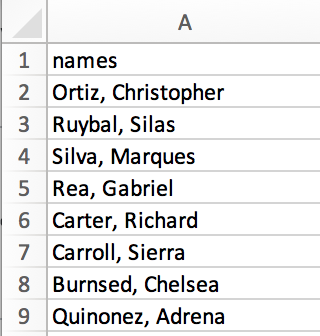
**Problem 1 (40 points)**

50 students registered for a mechanical design course. Their names are given in the course roster that can be accessed by the professor. The details of their names are given in the file titled “*roster.csv*”. The names are printed as “last name, first name” format. A snapshot of the roster data is given below



When the students attend a lecture, they enter the attendance for each class using an online form. The form records their first name and last name. The attendance data is given in the file titled “*attendance.csv*”. A snapshot of the student attendance data is given below



**Task**

1. Read the two data files in python
2. Write a code to aggregate attendance, i.e., how many lectures were attended by each student. If a student has not attended any lectures, set the value as zero. The code should be effective in case of multiple instances of the same first name or last name (or both). These values must be updated in the roster.
3. The output of the code should generate a roster data frame as given below.

|  |  |  |
| --- | --- | --- |
| **First name** | **Last name** | **Count** |
| Bob | Ross | 4 |
| Ron | Swanson | 0 |

**Problem 2**

From the “*wine\_data.csv*” answer the following questions using data wrangling functions from relevant packages

1. Write a code to calculate the frequency count of “*variety*” variable from the dataset. Display top 10 variety by count (**10 points**)
2. Write a code to calculate the average points by country (**10 points**)
3. Which province has the highest average price? (**10 points**)
4. Which province in the US has the highest average price? (**10 points**)
5. From the “*designation*” variable calculate the number of 20 year old wine (**20 points**)

**Submission Format**

1. Submit the solution in a python notebook in the following format.

# Homework 1

# Student name

# Section

# Problem 1-----------------------------------

# Problem 2-----------------------------------