Problem 1 (20 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

	A
1	names
2	Ortiz, Christopher
3	Ruybal, Silas
4	Silva, Marques
5	Rea, Gabriel
6	Carter, Richard
7	Carroll, Sierra
8	Burnsed, Chelsea
9	Quinonez, Adrena

The students enter the attendance for each class using an online form. The form records their first name and last name and send the data in csv format to the professor. The attendance data is given in the file titled "*attendance.csv*". A snapshot of the student attendance data is given below

	A	B
1	firstname	lastname
2	Matthew	Garza
3	Lashawn	King
4	Richard	Carter
5	Marques	Silva
6	Kyrie	Crow-Willard
7	Lashawn	King
8	Dominique	Lynch
9	Taneja	Jackson
10	Christian	Conway
11	Jose	Chacon
12	Michael	Kunimune
13	Jason	Baird
14	Aidan	Alexopoulos

Task

1. Read the two data files in R studio
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 same first name or last name (or both)
3. The output of the code should generate a data frame as given below.

First name	Last name	Count
Bob	Ross	4
Ron	Swanson	0

Problem 2 (20 points)

From the Boston crime dataset answer the following

	OFFENSE_CODE_GROUP	DISTRICT
1	Residential Burglary	B3
2	Drug Violation	B2
3	Motor Vehicle Accident Response	C11
4	Missing Person Reported	B2
5	Simple Assault	E13
6	Property Lost	*NA*
7	Police Service Incidents	D4
8	Investigate Person	B2
9	Motor Vehicle Accident Response	C11

	B3	B2	C11	E13
Residential Burglary	0	0	0	0
Drug Violation	0	0	0	0
r Vehicle Accident Response	0	0	0	0
Missing Person Reported	0	0	0	0
Simple Assault	0	0	0	0
Property Lost	0	0	0	0
Police Service Incidents	0	0	0	0
Investigate Person	0	0	0	0

Select only OFENSE_CODE_GROUP and DISTRICT from the Boston crime dataset (see image to the top left). Convert the two columns into a matrix (as shown in the top right image). Make sure that the matrix is populated. For example, if in district B3, Residential Burglary occurred 50 times then the cell corresponding to Residential Burglary and B3 in the matrix should be populated as 50. Compute this for all combinations of OFENSE_CODE_GROUP and DISTRICT. While creating the matrix make sure that *NA* from OFENSE_CODE_GROUP and DISTRICT is not included in either matrix rows or columns.

Datasource for problem 2:

https://data.boston.gov/dataset/crime-incident-reports-august-2015-to-date-source-new-system

Problem 3

From the "*wine_data.csv*" answer the following questions using data wrangling functions from relevant packages

a) Write a code to calculate the frequency count of "*variety*" variable from the dataset. Display top 10 variety by count (**10 points**)
b) Write a code to calculate the average points by country (**10 points**)
c) Which province has the highest average price? (**10 points**)
d) Which province in the US has the highest average price? (**10 points**)
e) From the "*designation*" variable calculate the number of 20 year old wine (**20 points**)