Prime Minister XX first term is coming to an end and he is therefore campaigning for a second term. To help PM in getting reelected he wants to introduce his signature piece of legislation in the form of a Tax Reform. PM wants to radically change the way that the people of his country pay taxes, but he is also painfully aware that a functional budget must still be possible.

Over the course of the past few weeks market researchers have been in the field to gather data. The data is assembled in three excel spreadsheets: Citizens, Taxes and Income.

As you are not the only analyst working on the data it is important that all plots and exports contain your employee number 409. All plots must have a title, a legend and, if relevant, titles on the x- and y-axis.

Citizens.xslx:

The spreadsheet Citizens.xslx contains personal information about 4922 citizens. The information is the following:

- The person's Social Security Number (SSN), which consists of 13 numbers in the format of DDMMYYYYXXXXX where DD is a two digit date of month of birth, MM is a two digit month of birth, YYYY is the year of birth and XXXXX is a personal number given at birth.
- The person's name
- The person's gender
- The region the person lives in. There are 15 regions identified by the numbers 1 to 15.

Taxes.xlsx: The spreadsheet Taxes.xlsx contains information about the tax level in the 15 regions. The current tax system works by charging one tax rate up to a limit (Tax 1) and a different tax rate above the limit (Tax 2). Each region has its own limit (Tax 2 limit). For example: A person makes 100.000,- per year in region 1. This person will pay 27% tax on the first 55.000,- and 43% tax on the remaining 45.000,-. The person therefore pays (0.27*55000)+(0.43*45000) = 38.500,- in taxes.

Income.xlsx: The spreadsheet Income.xlsx contains information about the interviewees personal income for the year 2016. The income is before tax.

Assignments:

1. Decide on a data structure to hold the information mentioned above. The data structure should be able to handle more citizens, taxes and incomes in the future. Please argue why you have selected the specific data structure.

2. Import the data from the excel spreadsheets into your data structure.

3. Create a function *calculateStatistics* that calculates or finds the following

    a) The total income for all the citizens

    b) The highest and lowest income

    c) The average income

    d) The average income for each gender

4. Create a function *calculateTaxes* that calculates the following

    a) The total tax payed in each region

    b) The average tax payed in each region

5. Create a function *plotStatistics* that creates the following plots

    a) The total tax payed in each region

    b) The total income in each region

    c) The number of citizens in each region

    d) The number of citizens paying only Tax 1 and the number of citizens that also pay Tax 2. This must be a pie chart.

Note: The function must be able to show all plots at the same time.

6. Create a function *lookUpCitizen* that takes a citizen SSN as an argument. The function must return the person's name, income and taxes payed.

7. Create a function *exportCitizens* that exports all the SSN, year of birth and income for all citizens. You should exclude the headers. Your function must handle the situation where the file already exists.

8. Create a function *importCitizens* that import the data exported in sub-assignment 7 into a new variable called importedCitizens. You function must handle the situation where the data structure is already populated with data.

9. Create a function *addCitizen* that allows the user to add a citizen to the system. Your function must take the information from the user either as arguments or using the input-function. Please argue why you choose to use the input method that you did.

10. Create a function *changeTax* that takes the region, "Tax 1" rate, "Tax 2 limit" and "Tax 2" rate as arguments. You function must calculate the consequences of the changes.

11. Divide the citizens into groups based on their age. You should have at least ten groups. Plot the average taxes payed for each group. Based on the plot alone, would you conclude that the citizen's average income rises as the citizens get older? Please argue your conclusion.

12. Assume that a "tax break" was created for the citizens over the age of 70, where they would only have to pay half the "Tax 2" rate (e.g. 45% becomes 22.5%). What would be the economic consequences of the "tax break" and which region would have the biggest change in tax income?