# CS571: Advanced Programming Techniques

Winter 2020

*Assignment #6*

**Due date: Tuesday, Feb 25 at 11:59pm**

**Population Analytics Using Python**

The United States Census Bureau, is the nation’s leading source of statistical information about the nation’s people, based on census data collected every 10 years, along with several other surveys, demographic analysis data and population estimations. One of the reports maintained by the Bureau is the annual estimates of US population statistics by single year of age and by sex. The latest is from 2017 and can be found at the following URL

<https://www2.census.gov/programs-surveys/popest/datasets/2010-2017/national/asrh/nc-est2017-agesex-res.csv>

This CSV file is also available, as agesex.csv, through BB Learn as part of the Assignment.

A description of the CSV file is available online at the following URL

<https://www2.census.gov/programs-surveys/popest/datasets/2010-2017/national/asrh/nc-est2017-agesex-res.pdf>

The SEX column contains numeric codes: 0 stands for the total, 1 for male, and 2 for female. The AGE column contains ages in completed years, but the special value 999 is a sum of the total population. The row with AGE 100 doesn't just represent 100-year-olds – it also includes those who are older than 100. That is why the numbers in that row are larger than in the row for the 99-year-olds. The rest of the columns contain estimates of the US population.

In this assignment we will compare the population estimations of 2010 and 2017 (just the columns POPESTIMATE2010 and POPESTIMATE2017).

The assignment is given in parts, but only one file a6.py should be submitted.

**Part 1 – Population Trends**

Write a Python program that reads the CSV file, compares the population estimates of every row for 2010 and 2017 (compares columns POPESTIMATE2010 and POPESTIMATE2017 only) and computes the difference in populations as well as the percentage of change (in 2 decimal points). Even though the program might perform the computations for every row, it should output only 3 rows: one for the total of both genders (SEX code 0), one for males (SEX code 1) and one for females (SEX code 2).

The output should look similar to this:

SEX AGE 2010 2017 CHANGE PERCENT\_CHANGE

0 999 … … …. ….

1 999 … … …. ….

2 999 … … …. ….

Which population changed more, the male population or the female population? The program should print a statement that tells us this information as well.

To convert a number num into a number with only two decimal points, use the function round(num,2).

**Part 2 – Overall proportions of males and females**

The program should be extended to show the total population of males and females in 2017. It should compute the proportion of males and females from the total population in 2017 and display an output similar to this:

SEX AGE 2017 Proportion

0 999 … 100%

1 999 … …

2 999 … …

In 2017, which population has the biggest proportion, males or females? The program should print a statement that tells us this information.