CHAPMAN University

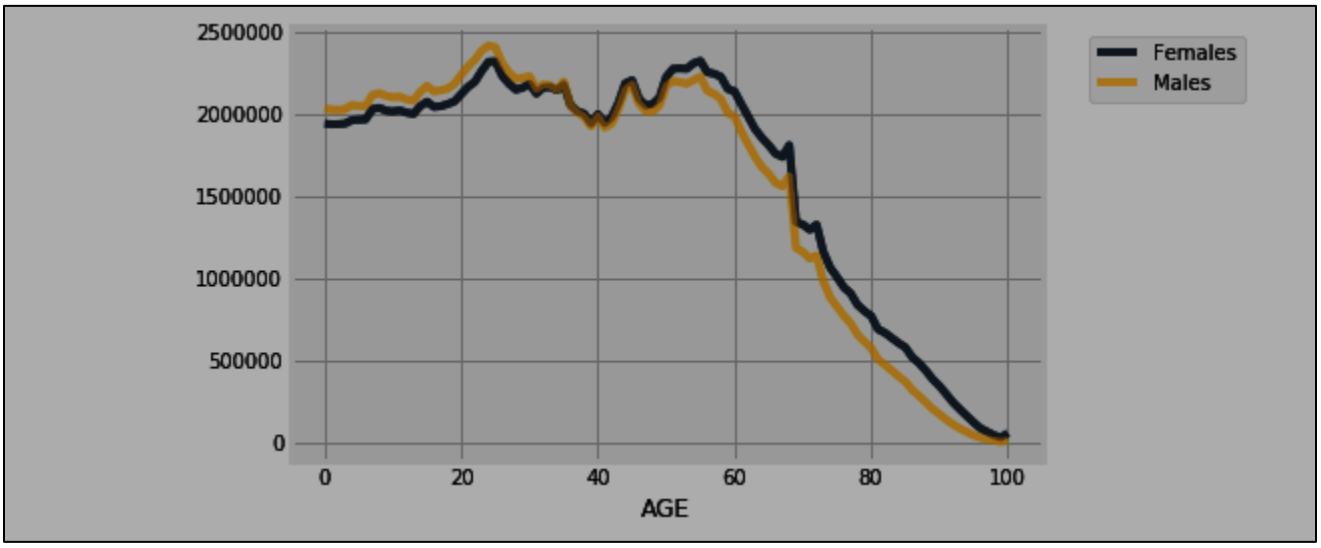
Department of Computational and Data Sciences
CS501 Introductory Computation for Scientists
Fall 2019
Class Project#1
Male/Female Life Expectancy

Date Given: Sep 18, 2019 Due Date: Oct 15, 2019

==========

Project goal: Using the US Census data prove that females live longer than males.

One of the ways to prove the above assertion is to generate the following plot.



Procedure to solve this problem.

US Government conducts census every 10 years. Census counts every person living in US regardless of their citizen status. Census data include age, gender and the home address of the person. Last census was done in 2010 and the next census will be next year - 2020. Every year in between census (2011 – 2019), government estimates the population growth and adds the estimated data to the table.

Census data can be downloaded using the following URL.

https://www2.census.gov/programs-surveys/

The URL of the census file needed for this project is as follows.

https://www2.census.gov/programs-surveys/popest/datasets/2010-2015/national/asrh/nc-est2015-agesex-res.csv

I have downloaded this file and posted on Blackboard. This file contains approximately 300 rows. The data coding as follows.

There are 3 codes for sex.

Sex Coding	Semantics
0	Male + Females
1	Male
2	Female

There are 102 codes for age.

Age Coding	Semantics
0	Age 0 – 1
1	Age 1 – 2
2	Age 2 - 3
…	
100	Age 100 or more
999	All ages 0 - 100

The population data is broken down by age from 0 – 100. People who are of age more than 100 are grouped in the last row of age 100.

Here is the sample data for the first 10 rows.

	A	B	C	D	E	F	G	H	I	J
1	SEX	AGE	CENSUS2010POP	ESTIMATESBASE2010	POPESTIMATE2010	POPESTIMATE2011	POPESTIMATE2012	POPESTIMATE2013	POPESTIMATE2014	POPESTIMATE2015
2	0	0	3944153	3944160	3951330	3963087	3926540	3931141	3949775	3978038
3	0	1	3978070	3978090	3957888	3966551	3977939	3942872	3949776	3968564
4	0	2	4096929	4096939	4090862	3971565	3980095	3992720	3959664	3966583
5	0	3	4119040	4119051	4111920	4102470	3983157	3992734	4007079	3974061
6	0	4	4063170	4063186	4077551	4122294	4112849	3994449	4005716	4020035
7	0	5	4056858	4056872	4064653	4087709	4132242	4123626	4006900	4018158
8	0	6	4066381	4066412	4073013	4074993	4097605	4142916	4135930	4019207
9	0	7	4030579	4030594	4043046	4083225	4084913	4108349	4155326	4148360
10	0	8	4046486	4046497	4025604	4053203	4093177	4095711	4120903	4167887
11	0	9	4148353	4148369	4125415	4035710	4063152	4104072	4108349	4133564
12	0	10	4172541	4172559	4187062	4135613	4045779	4074111	4116942	4121289

Interpretation of the first row of data.

- Sex = 0 (means males + females)
- Age = 0 (means babies from age 0 – 1)
- Population in 2010 = 3,944,153 (this is accurate)
- Population in 2011 = 3,963,087 (this is an estimate)
- Population in 2012 = 3,926,540 (this is an estimate)
- …
- Population 2015 = 3,978,038 (this is an estimate)

The following data is extracted from the above table.

Sex	Age	Population 2010 Census
0 (Males + Females)	0	3,944,153
1 (Males)	0	2,014,276
2 (Females)	0	1,929,877
		For verification: 2,014,276 (Males) + 1,929,877 (Females) = 3,944,153

The following data is extracted from the above table.

Sex	Age	Population 2010 Census
0 (Males + Females)	999	308,745,538
0 (Males + Females)	0 - 100	For verification: SUM of the 2010 Census = 308,745,538

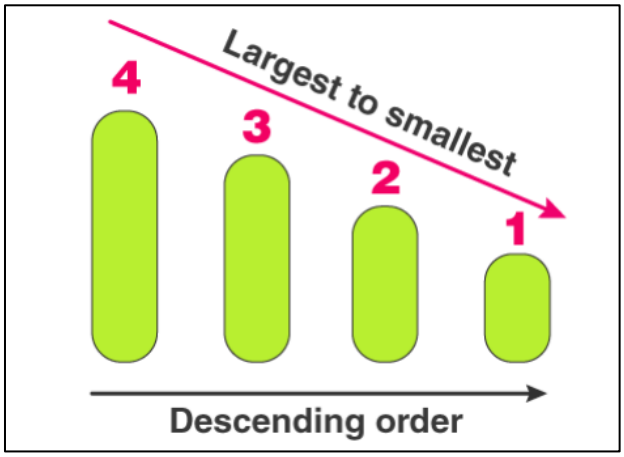
Although the data from year 2010 – 2015 is given, you should only use the following 2 columns.

- CENSUS2010POP
- POPESTIMATE2015

This class project involves writing the Python code to do the following using the dataset.

1. Compute the population **annual** percent growth from 2010 to 2015 for every age group (age code = 0 to 999) and gender (sex code = 0-2).

2. Sort the table (descending order) created in step number 1 displaying group with the highest population growth on the top.



3. Generate the following plot which proves the assertion that females live longer than males.