UNIX AWK ASSIGNMENT

You and your elbow-mate may choose one of four possible population data sets; Brooklyn's (BKLNpop.txt), New York City's (NYCpop.txt), New York State's (NYSpop.txt), or The United State's populations from 1900 to 2020. You will write and run an awk program to predict the populations from your data set for 2030, 2040, and 2050 censuses. You will use the code from tedWarYears.txt as a guide. The data sets you are given are not separated by commas, so you will NOT need

```
{FS = ","}
```

AWK defaults to blank spaces as field separators. Another hint, remember when you use your 'for' loop to make your predictions at the end, is that you are not predicting consecutive years but stepping by 10s, so your update section cannot be i++. When you are done writing your awk script in a text file, you will run it from the command line with this format

```
awk -f "filename.txt" inputfile.txt
```

There are only 3 fields in each data set, you will only need to use 2 to complete your assignment. Determine what is your x and y variable. Remember, it is the y variable you are trying to predict. The Linear regression Formula is

```
Y = a + bX
```

Answer the following questions. For credit, answer question 1. For an extra challenge, answer questions 2-4. 1. Using all available data points for your model (1900 - 2020) what does your model predict the population to be in 2030, 2040, and 2050? 2. Using all available data points for your model (1900 - 2010) what does your model predict the population to be in 2020? How much does your model differ from the real data? Can you calculate the percent difference? 3. Using just the previous 4 data points (1990-2020) what does your model predict the population to be in 2030, 2040, and 2050? How much does this answer differ from the results you got in question 1? Can you calculate the percent difference? 3. Using just the previous 4 data points (1980-2010) what does your model predict the population to be in 2000? How much does this answer differ from the results you got in question 2? Was it more accurate or less accurate? Can you calculate the percent difference?