Rodriguez_Felipe_DSC550_Week2Assignment

June 18, 2023

Week 2 Assignment

The data being used is world population data from a dataset on Kaggle by Ulrik Thyge Pedersen. This dataset contains four columns, country_code, country_name, year, and value. The dataset has population values from 1960 to 2021 and includes 265 countries. The questions that will be explored is to understand how the population of the United States has changed over time, if the population is still growing, and how it compares to Mexico and Canada.

```python
[1]: import pandas as pd
```

```python
[14]: data = pd.read_csv('population_by_country.csv')
      data.country_name.nunique()
```

```
[14]: 265
```

```python
[34]: df_us = data[data['country_name'] == 'United States']
```

```python
[67]: last_5_years = df_us.tail(5)
      last_5_years
```

```
[67]:       country_code   country_name   year         value
      15527           USA  United States   2017   325122128.0
      15528           USA  United States   2018   326838199.0
      15529           USA  United States   2019   328329953.0
      15530           USA  United States   2020   331501080.0
      15531           USA  United States   2021   331893745.0
```

```python
[86]: plt.plot(last_5_years.year, last_5_years.value)
      plt.title("Population in the last 5 years", fontsize=16)
      plt.xlabel("Year", fontsize=12)
      plt.ylabel("Population", fontsize=12)
      plt.locator_params(axis="both", integer=True, tight=True)
      plt.grid(True)
      plt.legend(['US'], loc=2)
```

```
[86]: <matplotlib.legend.Legend at 0x7f8a84735d60>
```
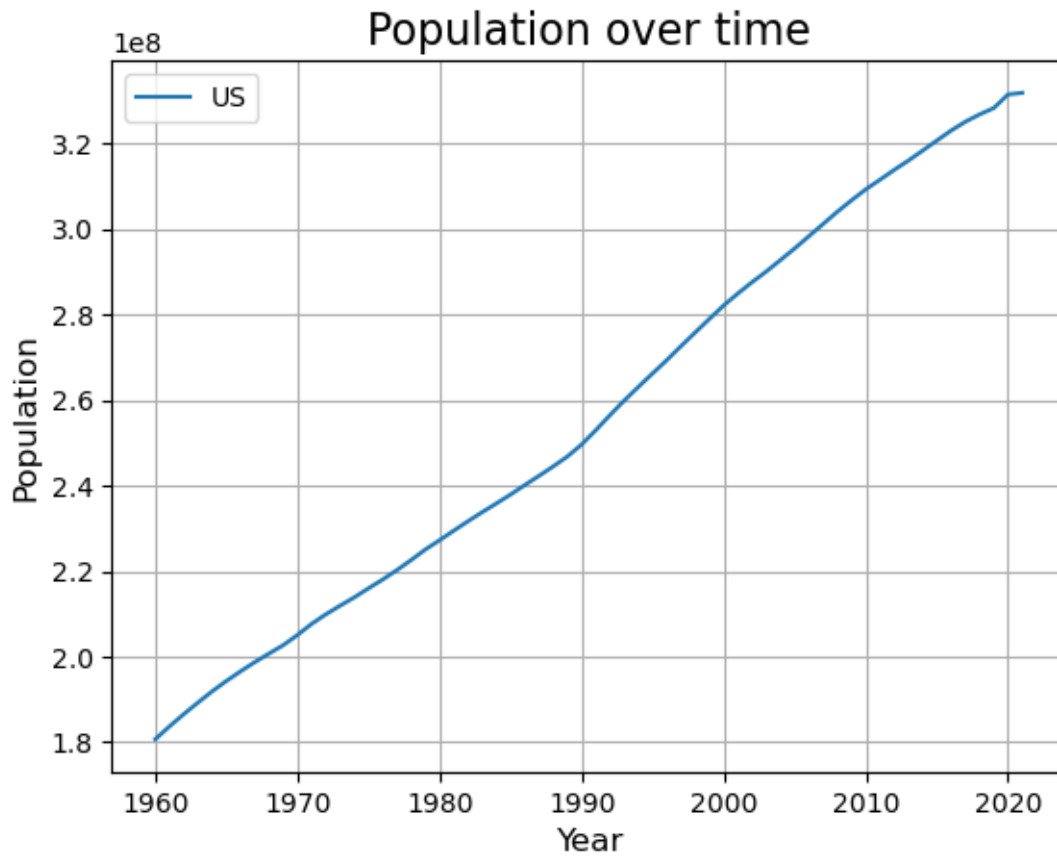
Population in the last 5 years
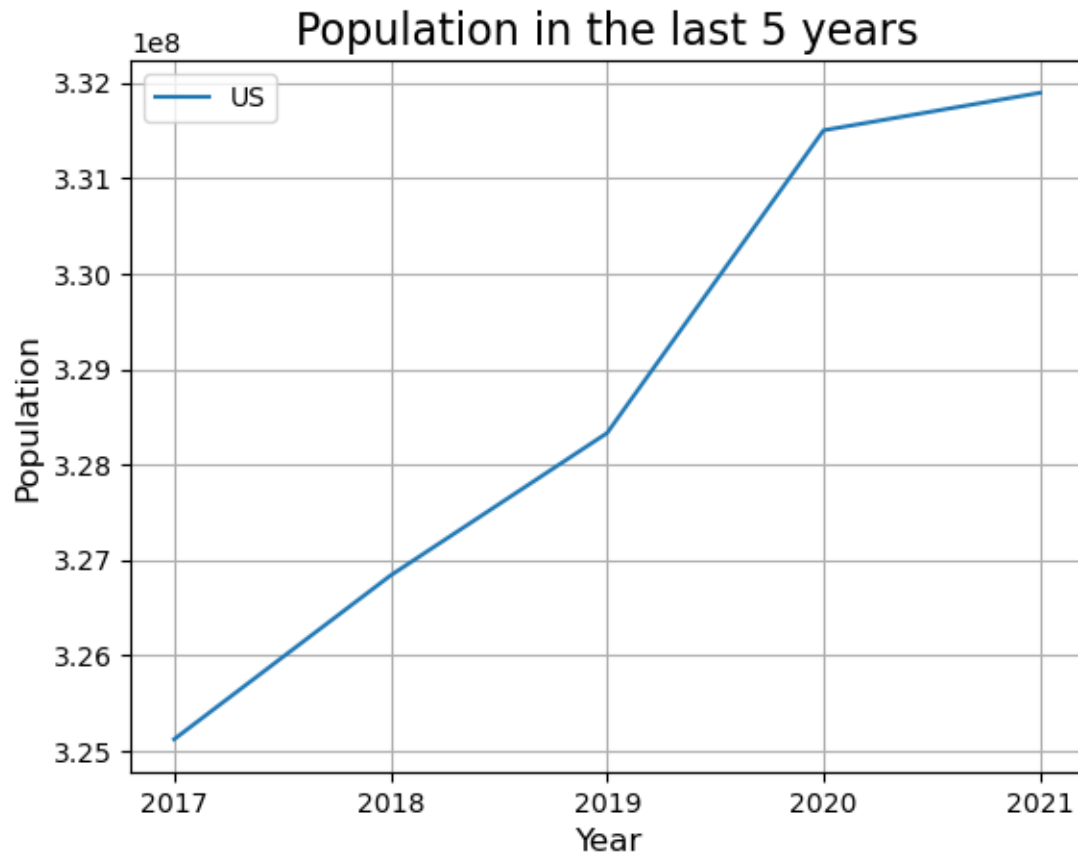
When looking at the most recent years of population data, it can be seen that the population has continued to increase. Even though there was a smaller increase from 2020 to 2021, this can potentially be due to COVID-19.

```
[15]: import matplotlib.pyplot as plt
```

```
[57]: plt.plot(df_us.year, df_us.value)
      plt.title("Population over time", fontsize=16)
      plt.xlabel("Year", fontsize=12)
      plt.ylabel("Population", fontsize=12)
      plt.grid(True)
      plt.legend(['US'], loc=2)
```

```
[57]: <matplotlib.legend.Legend at 0x7f8a50877940>
```
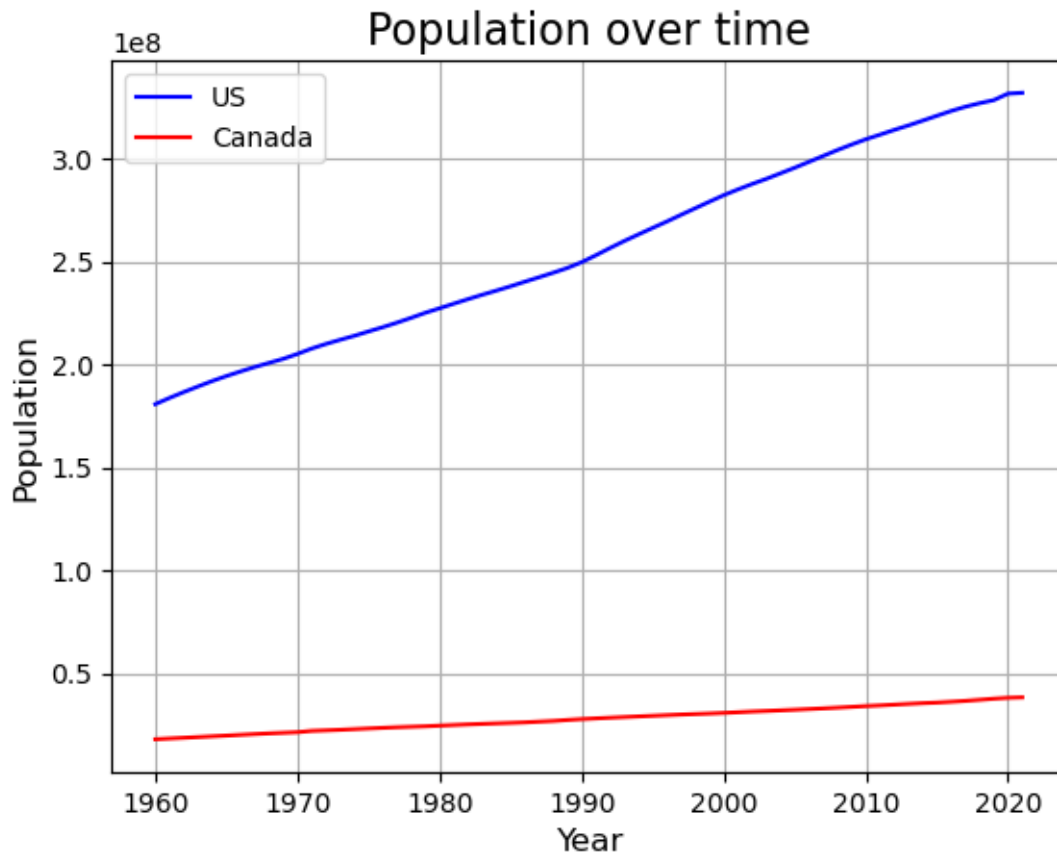
Population over time

The first plot is a visualization of the Population of the US over time. This graph gives insight on the population trend of over the last 50 years. Based on the graph, the population of the US has continued to grow.

```
[33]: df_cd = data[data['country_name'] == 'Canada']
```

```
[58]: plt.plot(df_us.year, df_us.value, color='blue')
      plt.plot(df_cd.year, df_cd.value, color='red')
      plt.title("Population over time", fontsize=16)
      plt.xlabel("Year", fontsize=12)
      plt.ylabel("Population", fontsize=12)
      plt.grid(True)
      plt.legend(['US', 'Canada'], loc=2)
```

```
[58]: <matplotlib.legend.Legend at 0x7f8a40146ca0>
```

Population over time

The next graph compares the US and Canda. The over the last 50 years, both countries have seen a growth of population, however, the rate of increase of the population of Canada is less due to the slope of population in the data.

```
[30]: df_mx = data[data['country_name'] == 'Mexico']
      df_mx
```

[30]:

	country_code	country_name	year	value
9486	MEX	Mexico	1960	36268055.0
9487	MEX	Mexico	1961	37439317.0
9488	MEX	Mexico	1962	38683283.0
9489	MEX	Mexico	1963	39982118.0
9490	MEX	Mexico	1964	41333878.0
...	...	...	...	...
9543	MEX	Mexico	2017	122839258.0
9544	MEX	Mexico	2018	124013861.0
9545	MEX	Mexico	2019	125085311.0
9546	MEX	Mexico	2020	125998302.0
9547	MEX	Mexico	2021	126705138.0

```
[62 rows x 4 columns]
```

```
[85]:  plt.plot(df_us.year, df_us.value, color='blue')
       plt.plot(df_mx.year, df_mx.value, color='green')
       plt.title("Population over time", fontsize=16)
       plt.xlabel("Year", fontsize=12)
       plt.ylabel("Population", fontsize=12)
       plt.grid(True)
       plt.legend(['US', 'Mexico'], loc=2)
```

```
[85]:  <matplotlib.legend.Legend at 0x7f8a600a8520>
```
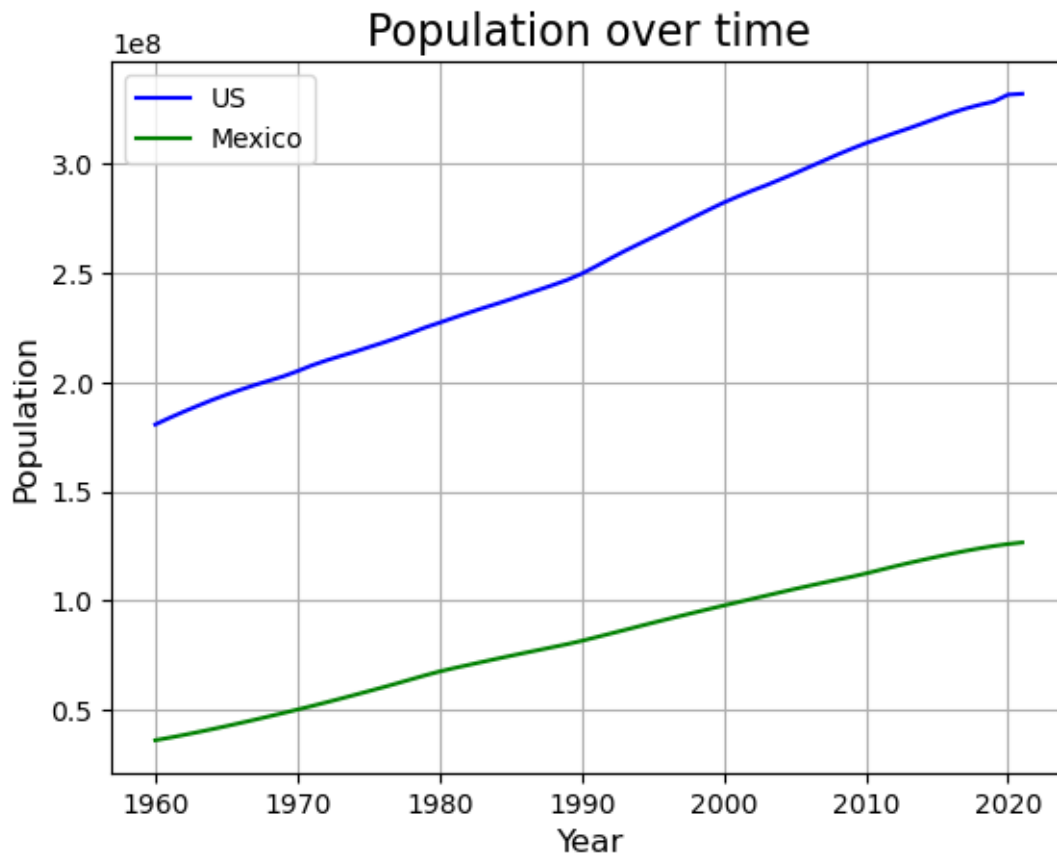

The next plot compares the populations of the US and Mexico. The population has also increased steadily in Mexico, and it seems to be at a slower rate as well.

```
[63]:  plt.plot(df_us.year, df_us.value, color='blue')
       plt.plot(df_cd.year, df_cd.value, color='red')
       plt.plot(df_mx.year, df_mx.value, color='green')
       plt.title("Population over time", fontsize=16)
       plt.xlabel("Year", fontsize=12)
```

```
plt.ylabel("Population", fontsize=12)
plt.grid(True)
plt.legend(['US', 'Canada', 'Mexico'], loc=2)
```

[63]: <matplotlib.legend.Legend at 0x7f8a8503b6d0>
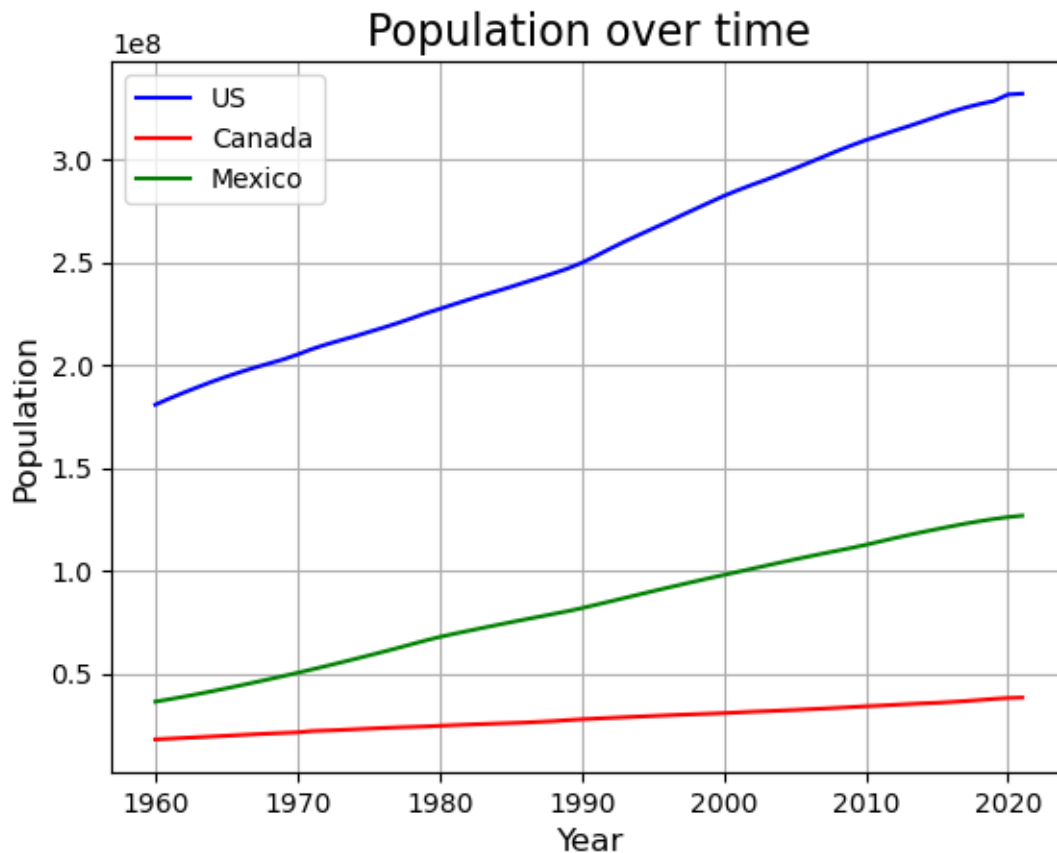
Lastly, this plot has all three countires displayed for a side by side comparison.

The graphs created gave insight into the population trends of the US, Canada, and Mexico. Overall, the US continues to grow at a higher rate out of all three and has a higher population than the other two countries. Canada is growing the slowest out of the three countries, but has not seen a negative trend yet. Mexico continues to grow at a consistent rate but will take some time to reach the population levels seen in the US. In the coming years, it will be interesting to analyze the changes in population for these three countries.