Topic 1 DQ 1

What are some examples of situations in which basic statistical concepts, such as population, sample, parameter, and statistic, could be useful for making decisions or drawing conclusions? How can Python be used to help analyze data in these situations?

Statistical data helps present factual evidence to help make decisions. Below is my example of how Python can be used to draw conclusions on where the US stacks up aginst other countires in terms of population and growth over the year. From this, we can draw a conclusion on our population growth compared to the other countries. Matplotlib made a simple bar graph to illustrate the correlation.

References:

Countries population by year 2020. (2020). Kaggle. [Data set]. https://www.kaggle.com/datasets/eng0mohamed0nabil/population-by-country-2020/discussion/160089

Rogel-Salazar, J. (2023). Statistics and data visualisation with python. CRC Press.

In [20]:
```
#import packages
import pandas as pd
import matplotlib.pyplot as plt
```

In [68]:
```
#reading csv
df = pd.read_csv(r'C:\Users\Matthew\Desktop\GCU Classes\DSC-510\population_by_country_2020.csv')
```

In [90]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 235 entries, 0 to 234
Data columns (total 11 columns):
 #   Column             Non-Null Count  Dtype  
---  ------             --------------  -----  
 0   Country            235 non-null    object 
 1   Population (2020)  235 non-null    int64  
 2   Yearly Change      235 non-null    object 
 3   Net Change         235 non-null    int64  
 4   Density (P/Km²)    235 non-null    int64  
 5   Land Area (Km²)    235 non-null    int64  
 6   Migrants (net)     201 non-null    float64
 7   Fert. Rate         235 non-null    object 
 8   Med. Age           235 non-null    object 
 9   Urban Pop %        235 non-null    object 
 10  World Share        235 non-null    object 
dtypes: float64(1), int64(4), object(6)
memory usage: 20.3+ KB
```

In [79]:
```
df
```

Out[79]:

	Country	Population (2020)	Yearly Change	Net Change	Density (P/Km²)	Land Area (Km²)	Migrants (net)	Fert. Rate	Med. Age	Urban Pop %	World Share
0	South Korea	51260395	0.09%	43877	527	97230	11731.0	1.1	44	82%	0.66%
1	Taiwan	23808164	0.18%	42899	673	35410	30001.0	1.2	42	79%	0.31%
2	Singapore	5840996	0.79%	46005	8358	700	27028.0	1.2	42	N.A.	0.08%
3	Puerto Rico	2874636	-2.47%	-72555	323	8870	-97986.0	1.2	44	N.A.	0.04%
4	Macao	647508	1.39%	8890	21645	30	5000.0	1.2	39	N.A.	0.01%
...	...	...	...	...	...	...	...	...	...	...	...
230	Montserrat	4991	0.06%	3	50	100	NaN	N.A.	N.A.	10%	0.00%
231	Falkland Islands	3458	3.05%	103	0	12170	NaN	N.A.	N.A.	66%	0.00%
232	Niue	1624	0.68%	11	6	260	NaN	N.A.	N.A.	46%	0.00%
233	Tokelau	1354	1.27%	17	136	10	NaN	N.A.	N.A.	0%	0.00%
234	Holy See	801	0.25%	2	2003	0	NaN	N.A.	N.A.	N.A.	0.00%

235 rows × 11 columns

In [113…

```
#Setting Parameters
Country_Specific = ['China','United States','Russia','Canada', 'India']
df[df['Country'].isin(Country_Specific)]
```

Out[113]:

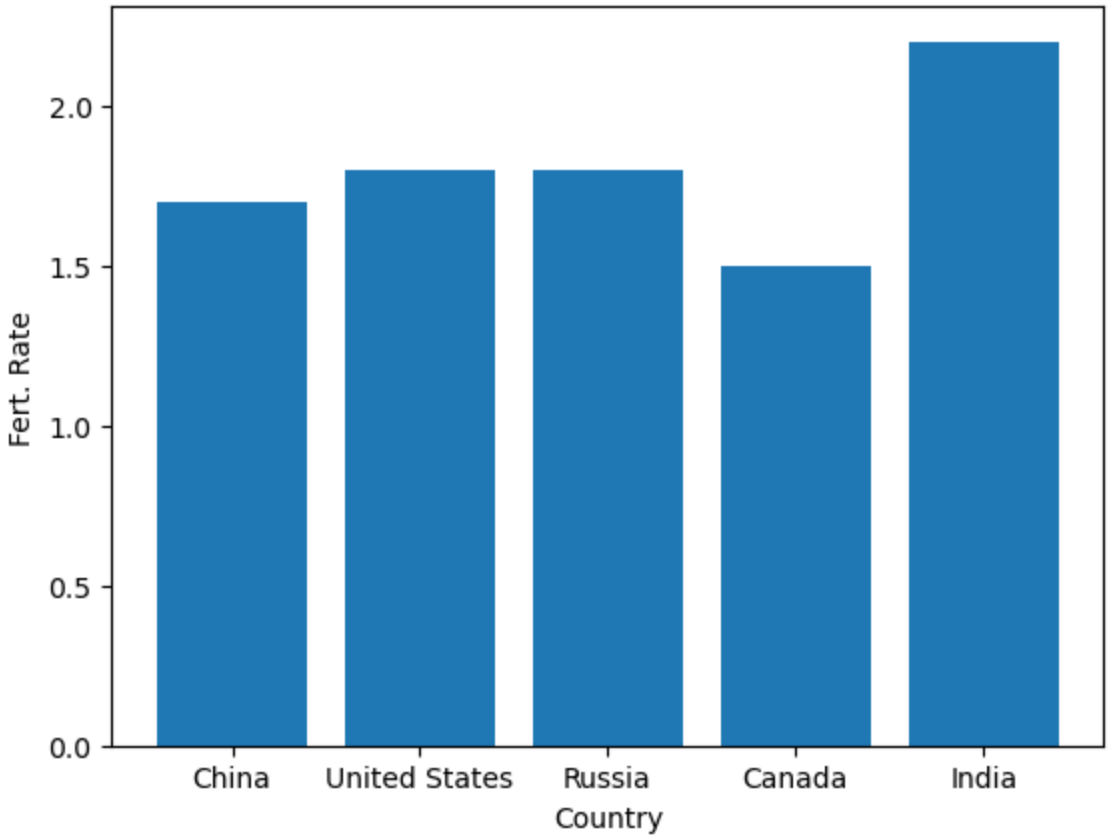
	Country	Population (2020)	Yearly Change	Net Change	Density (P/Km²)	Land Area (Km²)	Migrants (net)	Fert. Rate	Med. Age	Urban Pop %	World Share
21	Canada	37674770	0.89%	331107	4	9093510	242032.0	1.5	41	81%	0.48%
41	China	1438207241	0.39%	5540090	153	9388211	-348399.0	1.7	38	61%	18.47%
51	United States	330610570	0.59%	1937734	36	9147420	954806.0	1.8	38	83%	4.25%
52	Russia	145922010	0.04%	62206	9	16376870	182456.0	1.8	40	74%	1.87%
94	India	1377233523	0.99%	13586631	464	2973190	-532687.0	2.2	28	35%	17.70%

In [115…
```
Fertility_Rate = [1.7, 1.8, 1.8, 1.5, 2.2]
```

In [118…
```
#Plotting

plt.xlabel('Country')
plt.ylabel('Fert. Rate')
plt.bar(Country_Specific, Fertility_Rate)
```

Out[118]:
```
<BarContainer object of 5 artists>
```



In [ ]: