Stat 674 Quiz 1

Aaron Banlao

Instructions:

Complete the questions in the space provided. Render the Quarto Notebook to a .docx or .pdf file and submit your .qmd and your .docx or .pdf file through Canvas.

The quiz investigate the dataset *global_economy*. The dataset contains yearly historical economic data for all countries on Earth.

```
library(pacman)
p_load(tidyverse, fpp3)
```

Question 1

How many countries are there on Earth? (You should look this number up on Google or on Wikipedia.) How may Countries are there in the dataset? Why do the numbers differ?

Answer

There are 263 unique countries in the dataset. According to google, there are currently 193 countries. These countries are recognized as members of the United Nations, and the remainder of the countries are not recognized, hence the numbers differ in the dataset.

Provide your code here.

```
data(global_economy)
global_economy
```

```
# A tsibble: 15,150 x 9 [1Y]
# Key:       Country [263]
   Country     Code   Year         GDP Growth   CPI Imports Exports Population
   <fct>       <fct> <dbl>       <dbl>  <dbl> <dbl>   <dbl>   <dbl>      <dbl>
 1 Afghanistan AFG    1960  537777811.     NA    NA    7.02    4.13    8996351
 2 Afghanistan AFG    1961  548888896.     NA    NA    8.10    4.45    9166764
 3 Afghanistan AFG    1962  546666678.     NA    NA    9.35    4.88    9345868
 4 Afghanistan AFG    1963  751111191.     NA    NA   16.9     9.17    9533954
 5 Afghanistan AFG    1964  800000044.     NA    NA   18.1     8.89    9731361
 6 Afghanistan AFG    1965 1006666638.     NA    NA   21.4    11.3     9938414
 7 Afghanistan AFG    1966 1399999967.     NA    NA   18.6     8.57   10152331
 8 Afghanistan AFG    1967 1673333418.     NA    NA   14.2     6.77   10372630
 9 Afghanistan AFG    1968 1373333367.     NA    NA   15.2     8.90   10604346
10 Afghanistan AFG    1969 1408888922.     NA    NA   15.0    10.1    10854428
# ... with 15,140 more rows
```

```r
global_economy %>%
  distinct(Country) %>%
  count()
```

```
# A tibble: 1 x 1
      n
  <int>
1   263
```

Question 2

Create a new variable *GDP_per_capita*. Show the first few values of the new variable.

Provide your code here.

```r
global_economy <- global_economy %>%
  mutate(gdp_per_capita = GDP / Population)

global_economy %>%
  select(gdp_per_capita) %>%
  head()
```

```
# A tsibble: 6 x 3 [1Y]
# Key:       Country [1]
  gdp_per_capita  Year Country
           <dbl> <dbl> <fct>
1           59.8  1960 Afghanistan
2           59.9  1961 Afghanistan
3           58.5  1962 Afghanistan
4           78.8  1963 Afghanistan
5           82.2  1964 Afghanistan
6          101.   1965 Afghanistan
```

Question 3

Plot the time series data for *Population* for each of these countries: United States, Brasil, Canada, Mexico, Russia, Israel, and Japan. What do you notice about the population of Russian and Japan?

Answer

The population for Russia and Japan both reach a peak with their population that eventually declines. All the other countries have an increasing population trend.

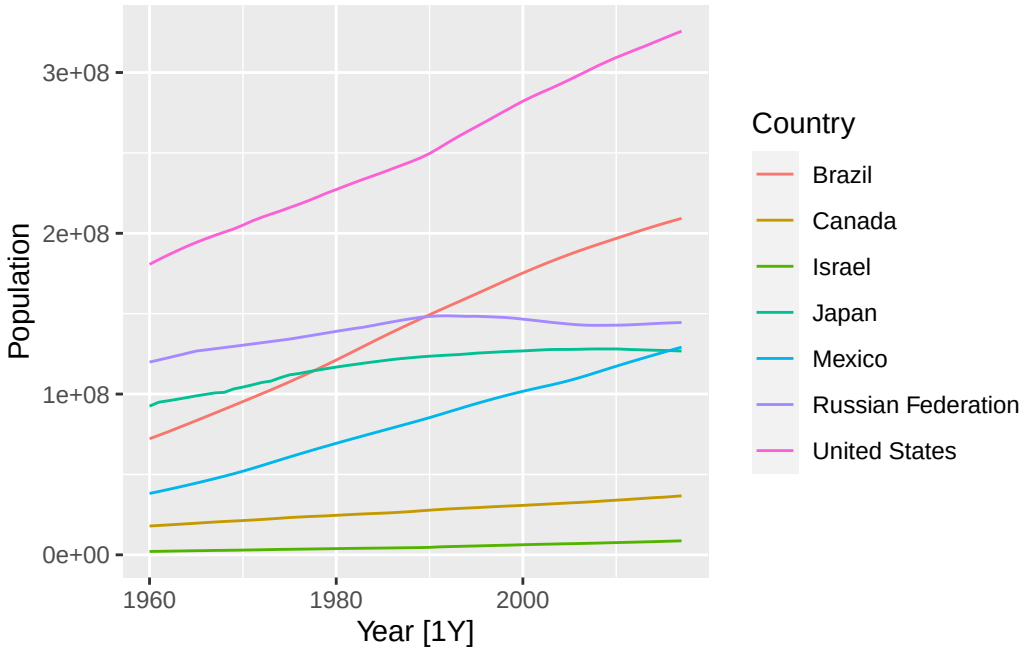
Provide your code here.

```
subset <- global_economy %>%
  filter(Country %in% c("United States", "Brazil", "Canada", "Mexico", "Russian Federation

head(subset)
```

```
# A tsibble: 6 x 10 [1Y]
# Key:       Country [1]
  Country Code   Year          GDP Growth   CPI Imports Exports Popula~1 gdp_p~2
  <fct>   <fct> <dbl>        <dbl>  <dbl> <dbl>   <dbl>   <dbl>    <dbl>   <dbl>
1 Brazil  BRA    1960 15165569913. NA        NA    7.12    7.06 72207554    210.
2 Brazil  BRA    1961 15236854859. 10.3      NA    7.34    7.28 74351763    205.
3 Brazil  BRA    1962 19926293839.  5.22     NA    5.19    3.87 76573248    260.
4 Brazil  BRA    1963 23021477292.  0.875    NA    9.11    9.04 78854019    292.
5 Brazil  BRA    1964 21211892260.  3.49     NA    5.68    6.39 81168654    261.
6 Brazil  BRA    1965 21790035117.  3.05     NA    5.56    7.74 83498020    261.
# ... with abbreviated variable names 1: Population, 2: gdp_per_capita
```

```
subset %>%
  autoplot(Population)
```



Question 4

Plot the time series data for *GDP* for each of these countries: United States, Brasil, Canada, Mexico, Russia, Israel, and Japan. How does the GDP of Japan compare to the GDP of the United States?

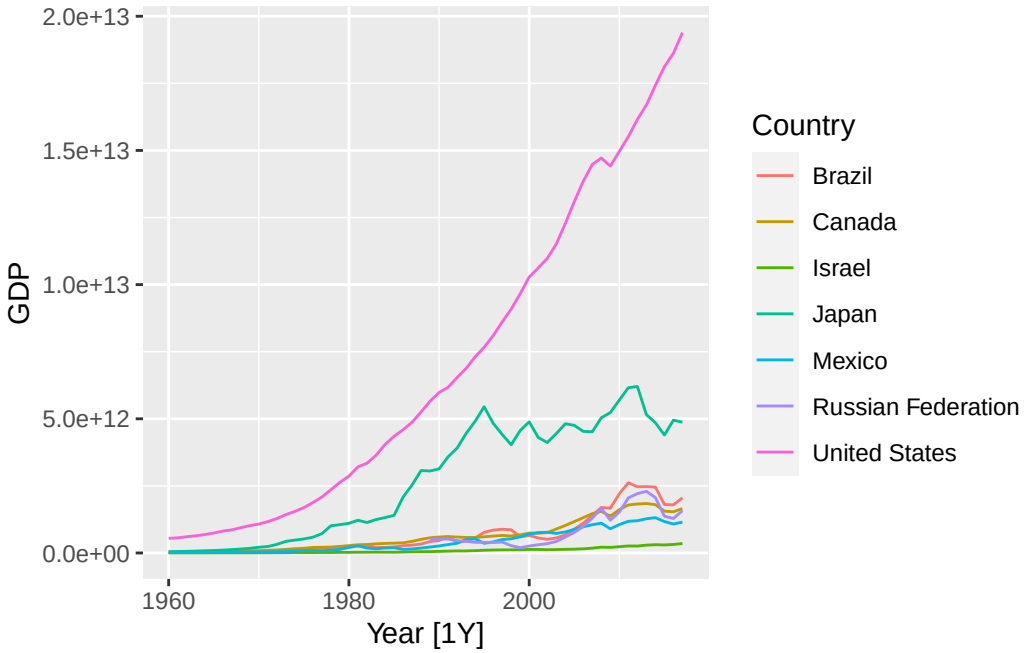
Answer

The GDP for the United States has an exponential looking upward trend, while Japan's GDP is much lower and fluctuates more than the other countries.

Provide your code here.

```
subset %>%
  autoplot(GDP)
```

```
Warning: Removed 29 rows containing missing values (`geom_line()`).
```



Question 5

Plot the time series data for *GDP_per_capita* for each of these countries: United States, Brasil, Canada, Mexico, Russia, Israel, and Japan. How does the GDP per capita differ for Russia, Mexico and Brasil?

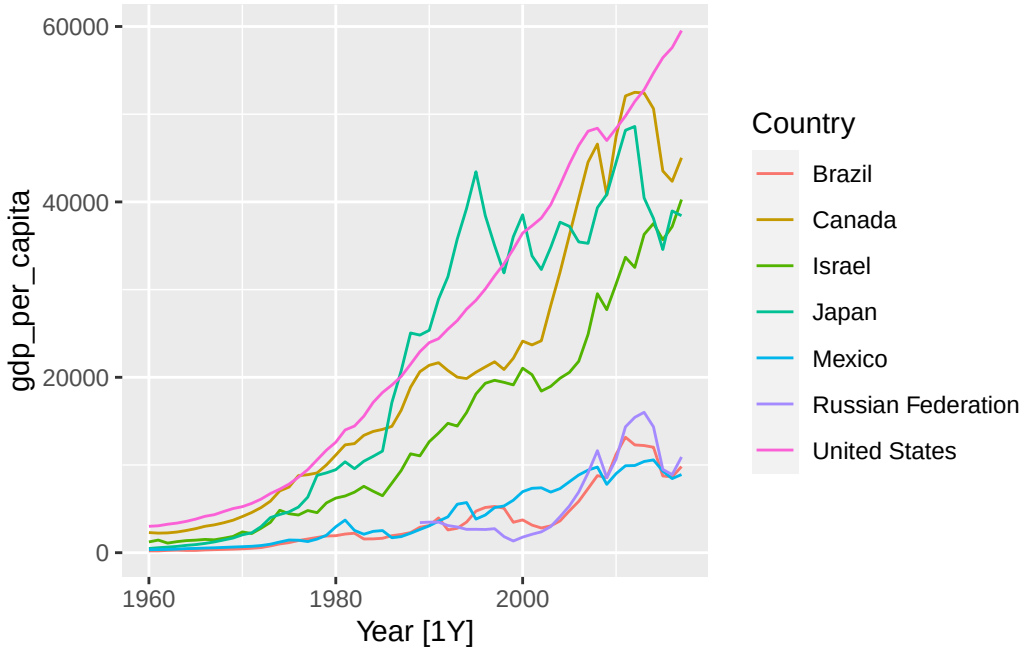
Answer

The GDP per capita for Russia, Mexico, and Brazil is mostly constant until 2000 and then finally has an upward trend with fluctuations.

Provide your code here.

```
subset %>%
  autoplot(gdp_per_capita)
```

```
Warning: Removed 29 rows containing missing values (`geom_line()`).
```



Question 6

Remake all of your plots including China. How does China compare to the United States in each plot?

Answer

China's population is substantially bigger than the United States which makes sense since China is a big country. The United States has a higher GDP than China but they both are similar in shape. The United States also has a higher GDP per capita but China's shape differs with more fluctuations.

Provide your code here.

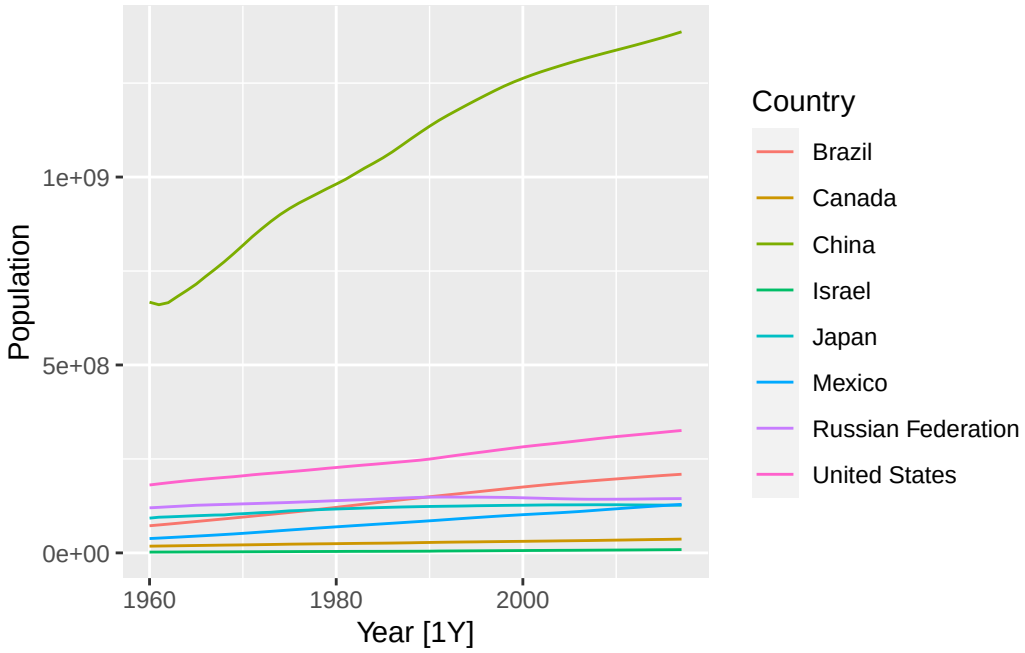
```
subset2 <- global_economy %>%
  filter(Country %in% c("United States", "Brazil", "Canada", "Mexico", "Russian Federation

subset2
```

```
# A tsibble: 464 x 10 [1Y]
# Key:       Country [8]
```

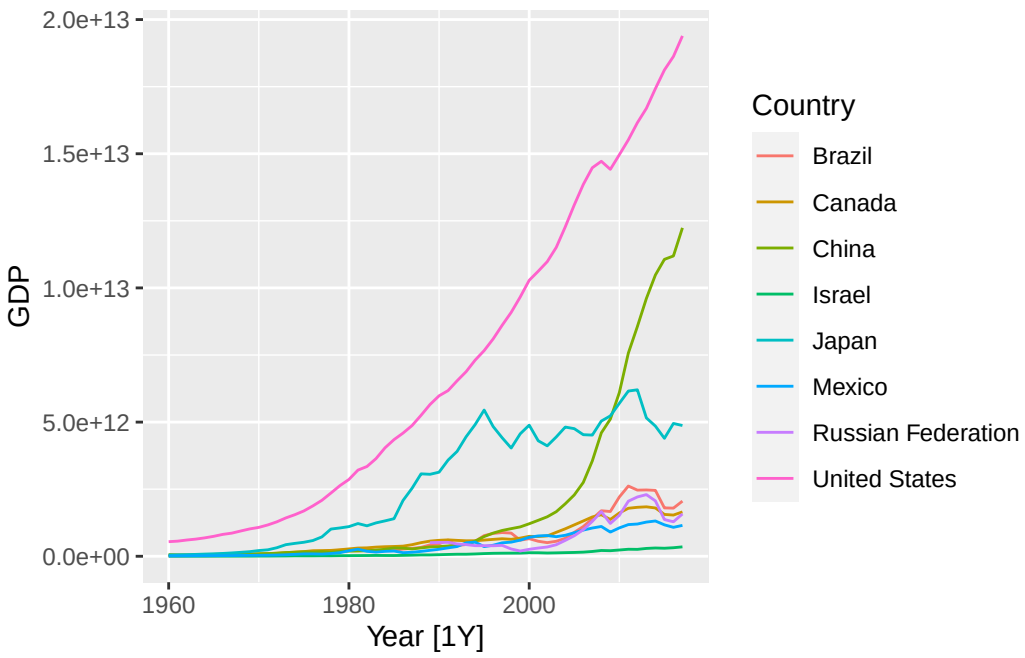
```
   Country Code   Year          GDP Growth   CPI Imports Exports Popul~1 gdp_p~2
   <fct>   <fct> <dbl>        <dbl>  <dbl> <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
 1 Brazil  BRA    1960 15165569913. NA        NA    7.12    7.06  7.22e7    210.
 2 Brazil  BRA    1961 15236854859. 10.3      NA    7.34    7.28  7.44e7    205.
 3 Brazil  BRA    1962 19926293839.  5.22     NA    5.19    3.87  7.66e7    260.
 4 Brazil  BRA    1963 23021477292.  0.875    NA    9.11    9.04  7.89e7    292.
 5 Brazil  BRA    1964 21211892260.  3.49     NA    5.68    6.39  8.12e7    261.
 6 Brazil  BRA    1965 21790035117.  3.05     NA    5.56    7.74  8.35e7    261.
 7 Brazil  BRA    1966 27062716578.  4.15     NA    5.99    6.82  8.58e7    315.
 8 Brazil  BRA    1967 30591834054.  4.92     NA    5.77    5.77  8.82e7    347.
 9 Brazil  BRA    1968 33875881876. 11.4      NA    6.61    6.00  9.06e7    374.
10 Brazil  BRA    1969 37458898244.  9.74     NA    6.55    6.62  9.29e7    403.
# ... with 454 more rows, and abbreviated variable names 1: Population,
#   2: gdp_per_capita
```

```
subset2 %>%
  autoplot(Population)
```



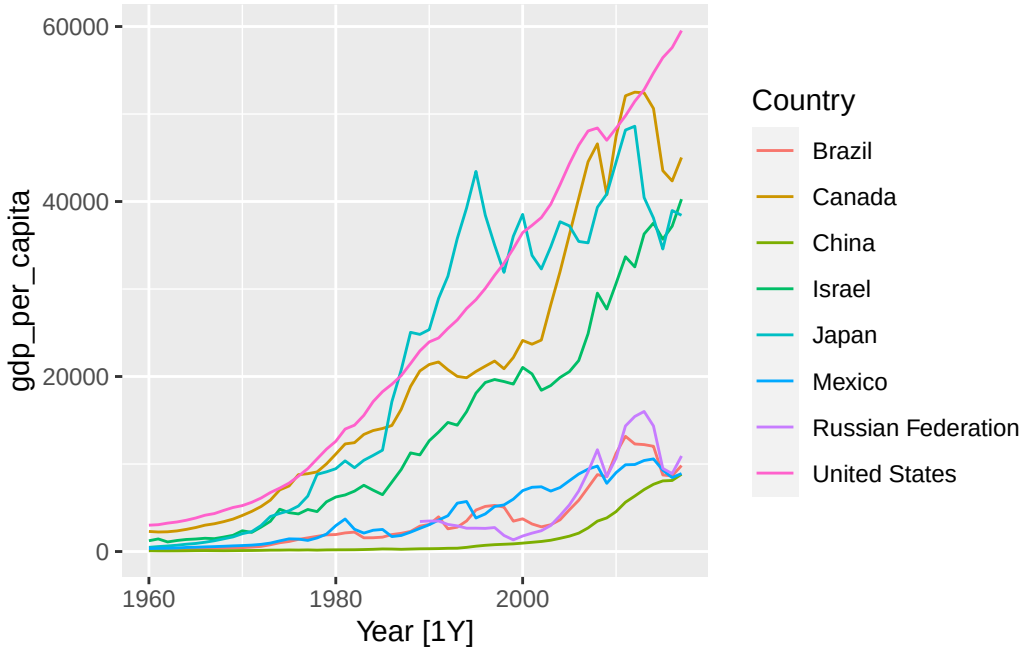

```
subset2 %>%
  autoplot(GDP)
```

```
Warning: Removed 29 rows containing missing values (`geom_line()`).
```




```
subset2 %>%
  autoplot(gdp_per_capita)
```

```
Warning: Removed 29 rows containing missing values (`geom_line()`).
```



Question 7

Does it make sense to run a Seasonal Decomposition with these data? Why or why not, explain.

Answer

It would not make sense to do a seasonal decomposition with this dataset because the time interval is yearly. Therefore it cannot capture the seasonality occuring in between months, days, etc.

Question 8

R questions.

a) Explain what a *tsibble* is?

A tsibble is a dataframe that uses a time interval such as year, month, day, etc. as an index.

b) Explain what a *mable* is?

A mable, or model table, is an object that gets returned when a model is applied to each of the key variables of a tsibble.

c) Explain what a *fable* is?

A fable, or forecast table, is an object that returns a new column that forecasts the interval of possible values for the next specified years, months, or days, etc.