# <Week 6 Mini Project>

<Yiwen Zhao>

# Dataset(s)

Which dataset did you use of the following:

- **World Development Indicators Dataset**

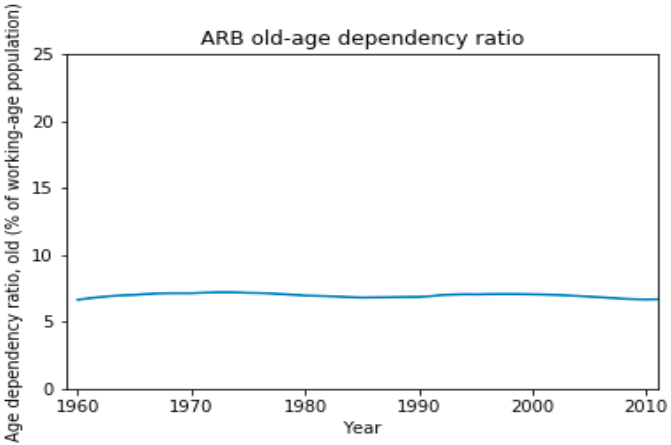
# Motivation

Under the trend of aging population, labor supply is declining, social production capacity is insufficient, and economic growth is declining. Second, aging will change the direction of social resources allocation and have an impact on economic growth. Generally speaking, the proportion of the elderly in an aging society is increasing, and the elderly belong to the "consumption-type" population in the economy and do not create output

# Research Question(s)

- **ARB old-age dependency ratio over the years**
- **What is the relationship between GDP and old-age dependency ratio in ARB?**

# Findings 1




## Yiwen Zhao Mini Project

**Question : ARB old-age dependency ratio over the years, and what is the relationship between GDP and old-age dependency ratio in ARB?**

In [2]:
```
import pandas as pd
import numpy as np
import random
import matplotlib.pyplot as plt
```

**let's inport data first**

In [3]:
```
data = data = pd.read_csv('./Indicators.csv')
```

**let's create a new dataframe "m" that contains only 'old-age dependency ratio' for the Country ARB**

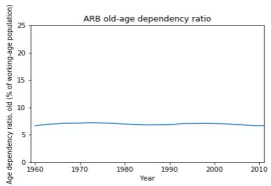
In [4]:
```
hist_old = 'Age dependency ratio, old \('
hist_contr = 'ARB'

m1 = data['IndicatorName'].str.contains(hist_old)
m2 = data['CountryCode'].str.contains(hist_contr)

# create a new df to meet above two requirements

m = data[m1 & m2]
m.head(3)
```

Out[4]:

| | CountryName | CountryCode | IndicatorName | IndicatorCode | Year | Value |
|---|---|---|---|---|---|---|
| 2 | Arab World | ARB | Age dependency ratio, old (% of working-age po... | SP.POP.DPND.OL | 1960 | 6.634579 |
| 23190 | Arab World | ARB | Age dependency ratio, old (% of working-age po... | SP.POP.DPND.OL | 1961 | 6.746488 |
| 49807 | Arab World | ARB | Age dependency ratio, old (% of working-age po... | SP.POP.DPND.OL | 1962 | 6.840777 |

**Creating a line chart to illustrate ARB old-age dependency ratio**

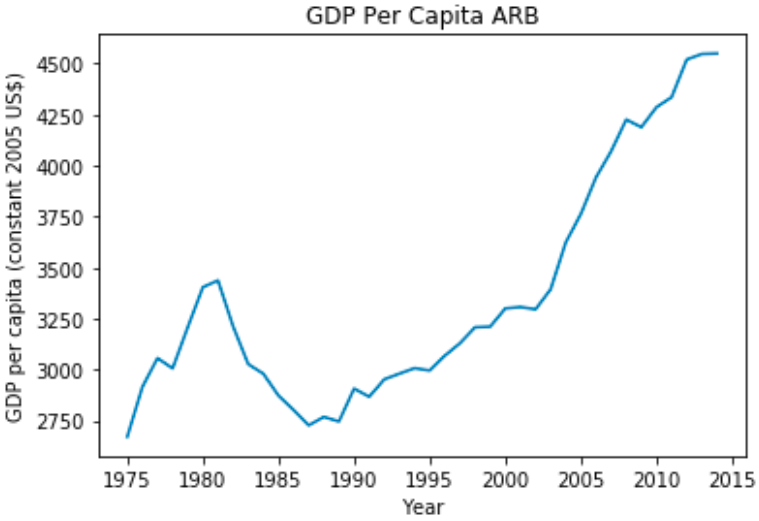
In [7]:
```
# switch to a line plot
plt.plot(m['Year'].values, m['Value'].values)

# Label the axes
plt.xlabel('Year')
# plt.ylabel(stage['IndicatorName'].iloc[0])
plt.ylabel(m['IndicatorName'].iloc[0])
#label the figure
plt.title('ARB old-age dependency ratio')

# to make more honest, start they y axis at 0
plt.axis([1959, 2011,0,25])

plt.show()
```




It's clear to see that there was almost no change of old-age dependency ratio between 1960 and 2010 in the ARB

**we can also see if there is any relationship between old-age dependency ratio and GDP**

# Findings 2




**Let's create a new dataframe 'ARB_gdp' which is the ARB GDP over the years.**

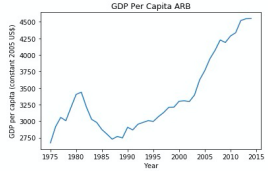
In [10]:
```
# select GDP Per capita emissions for the United States
hist_indicator = 'GDP per capita \(constant 2005'
hist_country = 'ARB'

mask1 = data['IndicatorName'].str.contains(hist_indicator)
mask2 = data['CountryCode'].str.contains(hist_country)

# stage is just those indicators matching the USA for country code and CO2 emissions over
ARB_gdp = data[mask1 & mask2]
```

In [11]:
```
ARB_gdp.head(2)
```

Out[11]:

| | CountryName | CountryCode | IndicatorName | IndicatorCode | Year | Value |
|---|---|---|---|---|---|---|
| 637939 | Arab World | ARB | GDP per capita (constant 2005 US$) | NY.GDP.PCAP.KD | 1975 | 2671.447438 |
| 710493 | Arab World | ARB | GDP per capita (constant 2005 US$) | NY.GDP.PCAP.KD | 1976 | 2915.579098 |

In [13]:
```
# switch to a line plot
plt.plot(ARB_gdp['Year'].values, ARB_gdp['Value'].values)

# Label the axes
plt.xlabel('Year')
plt.ylabel(ARB_gdp['IndicatorName'].iloc[0])

#label the figure
plt.title('GDP Per Capita ARB')

# to make more honest, start they y axis at 0
#plt.axis([1959, 2011,0,25])

plt.show()
```




**ScatterPlot for comparing GDP against old-age dependency ratio emissions (per capita)**

First, we'll need to make sure we're looking at the same time frames

In [14]:
```
print("GDP Min Year = ", ARB_gdp['Year'].min(), "max: ", ARB_gdp['Year'].max())
print("Old ratio Min Year = ", m['Year'].min(), "max: ", m['Year'].max())
```

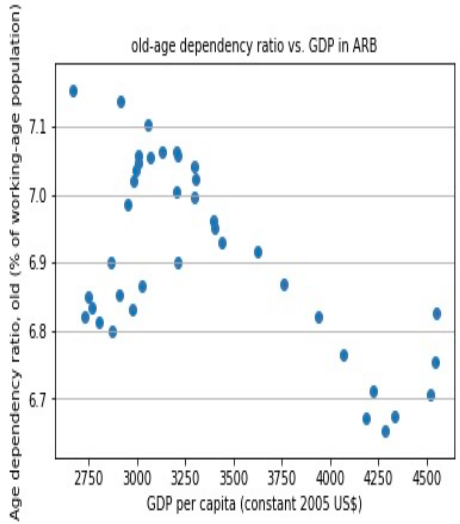
```
GDP Min Year =  1975 max:  2014
Old ratio Min Year =  1960 max:  2014
```

In [18]:
```
old_ratio_trunc = m[m['Year'] >= 1975]

print("GDP Min Year = ", ARB_gdp['Year'].min(), "max: ", ARB_gdp['Year'].max())
print("Old ratio Min Year = ", old_ratio_trunc['Year'].min(), "max: ", old_ratio_trunc['Y
```

```
GDP Min Year =  1975 max:  2014
Old ratio Min Year =  1975 max:  2014
```

In [19]:
```
%matplotlib inline
import matplotlib.pyplot as plt

fig, axis = plt.subplots()
# Grid lines, Xticks, Xlabel, Ylabel

axis.yaxis.grid(True)
axis.set_title('old-age dependency ratio vs. GDP in ARB',fontsize=10)
axis.set_xlabel(ARB_gdp['IndicatorName'].iloc[0],fontsize=10)
axis.set_ylabel(old_ratio_trunc['IndicatorName'].iloc[0],fontsize=10)

X = ARB_gdp['Value']
Y = old_ratio_trunc['Value']

axis.scatter(X, Y)
plt.show()
```

# Findings 3




Under the trend of aging population, labor supply is declining, social production capacity is insufficient, and economic growth is declining. Second, aging will change the direction of social resources allocation and have an impact on economic growth. Generally speaking, the proportion of the elderly in an aging society is increasing, and the elderly belong to the "consumption-type" population in the economy and do not create output

# Acknowledgements

I have one friend who came from Saudi Arabia, I showed my findings to him and he said that my result is close to the real life.

# References

Data Source: https://www.kaggle.com/worldbank/world-development-indicators