

Assignment-5

Iris Wu

2001-02-25

```
!pip install wbgapi
!pip install numpy
```

```
Requirement already satisfied: wbgapi in /Users/qwuiris/anaconda3/lib/python3.11/site-packages
Requirement already satisfied: requests in /Users/qwuiris/anaconda3/lib/python3.11/site-packages
Requirement already satisfied: PyYAML in /Users/qwuiris/anaconda3/lib/python3.11/site-packages
Requirement already satisfied: tabulate in /Users/qwuiris/anaconda3/lib/python3.11/site-packages
Requirement already satisfied: charset-normalizer<4,>=2 in /Users/qwuiris/anaconda3/lib/python3.11/site-packages
Requirement already satisfied: idna<4,>=2.5 in /Users/qwuiris/anaconda3/lib/python3.11/site-packages
Requirement already satisfied: urllib3<3,>=1.21.1 in /Users/qwuiris/anaconda3/lib/python3.11/site-packages
Requirement already satisfied: certifi>=2017.4.17 in /Users/qwuiris/anaconda3/lib/python3.11/site-packages
Requirement already satisfied: numpy in /Users/qwuiris/anaconda3/lib/python3.11/site-packages
```

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import wbgapi as wb
```

```
# Define the indicators to download
indicators = {
    'gdp_per_capita': 'NY.GDP.PCAP.CD',
    'gdp_growth_rate': 'NY.GDP.MKTP.KD.ZG',
    'inflation_rate': 'FP.CPI.TOTL.ZG',
    'unemployment_rate': 'SL.UEM.TOTL.ZS',
    'total_population': 'SP.POP.TOTL',
    'life_expectancy': 'SP.DYN.LE00.IN',
    'adult_literacy_rate': 'SE.ADT.LITR.ZS',
    'income_inequality': 'SI.POV.GINI',
    'health_expenditure_gdp_share': 'SH.XPD.CHEX.GD.ZS',
    'measles_immunisation_rate': 'SH.IMM.MEAS',
```

```

    'education_expenditure_gdp_share': 'SE.XPD.TOTL.GD.ZS',
    'primary_school_enrolment_rate': 'SE.PRM.ENRR',
    'exports_gdp_share': 'NE.EXP.GNFS.ZS'
}

# Get the list of country codes for the "World" region
country_codes = wb.region.members('WLD')

# Download data for countries only in 2022
df = wb.data.DataFrame(indicators.values(), economy=country_codes, time=2022, skipBlanks=True)

# Delete the 'economy' column
df = df.drop(columns=['economy'], errors='ignore')

# Create a reversed dictionary mapping indicator codes to names
# Rename the columns and convert all names to lowercase
df.rename(columns=lambda x: {v: k for k, v in indicators.items()}.get(x, x).lower(), inplace=True)

# Sort 'country' in ascending order
df = df.sort_values('country', ascending=True)

# Reset the index after sorting
df = df.reset_index(drop=True)

# Display the number of rows and columns
print(df.shape)

# Display the first few rows of the data
print(df.head(3))

# Save the data to a CSV file
df.to_csv('wdi.csv', index=False)

```

(217, 14)

	country	inflation_rate	exports_gdp_share	gdp_growth_rate \
0	Afghanistan	NaN	18.380042	-6.240172
1	Albania	6.725203	37.197085	4.826688
2	Algeria	9.265516	30.808979	3.600000

	gdp_per_capita	adult_literacy_rate	primary_school_enrolment_rate \
0	357.261153	NaN	NaN
1	6846.426143	98.5	96.371231

```

2      4961.552577      NaN      108.343933

      education_expenditure_gdp_share  measles_immunisation_rate  \
0      NaN      56.0
1      2.744330      86.0
2      4.749247      79.0

      health_expenditure_gdp_share  income_inequality  unemployment_rate  \
0      NaN      NaN      14.100
1      NaN      NaN      10.137
2      NaN      NaN      12.346

      life_expectancy  total_population
0      62.879      40578842.0
1      76.833      2777689.0
2      77.129      45477389.0

```

Exploratory Data Analysis

```

df_selected = df[['country', 'gdp_per_capita', 'unemployment_rate', 'education_expenditure_gdp_share']]
# Summary statistics
df_selected.describe()

```

/Users/qwuiris/anaconda3/lib/python3.11/site-packages/IPython/core/formatters.py:344: FutureWarning: DataFrame.describe() is deprecated and will be removed in a future version. Use DataFrame.describe_stats_only() to only get summary statistics or DataFrame.describe_transpose() to get summary statistics in a transposed DataFrame.

In future versions `DataFrame.to_latex` is expected to utilise the base implementation of `Series.to_latex`.

	gdp_per_capita	unemployment_rate	education_expenditure_gdp_share
count	207.000000	186.000000	137.000000
mean	20520.336828	7.227344	4.164884
std	30640.741594	5.844462	1.771027
min	250.634225	0.130000	0.348517
25%	2599.752468	3.478000	2.951592
50%	7606.237525	5.334000	3.938396
75%	27542.145523	9.261750	4.959176
max	226052.001905	35.359000	10.703345

The summary statistics provide insights into the range, mean, and distribution of the selected indicators:

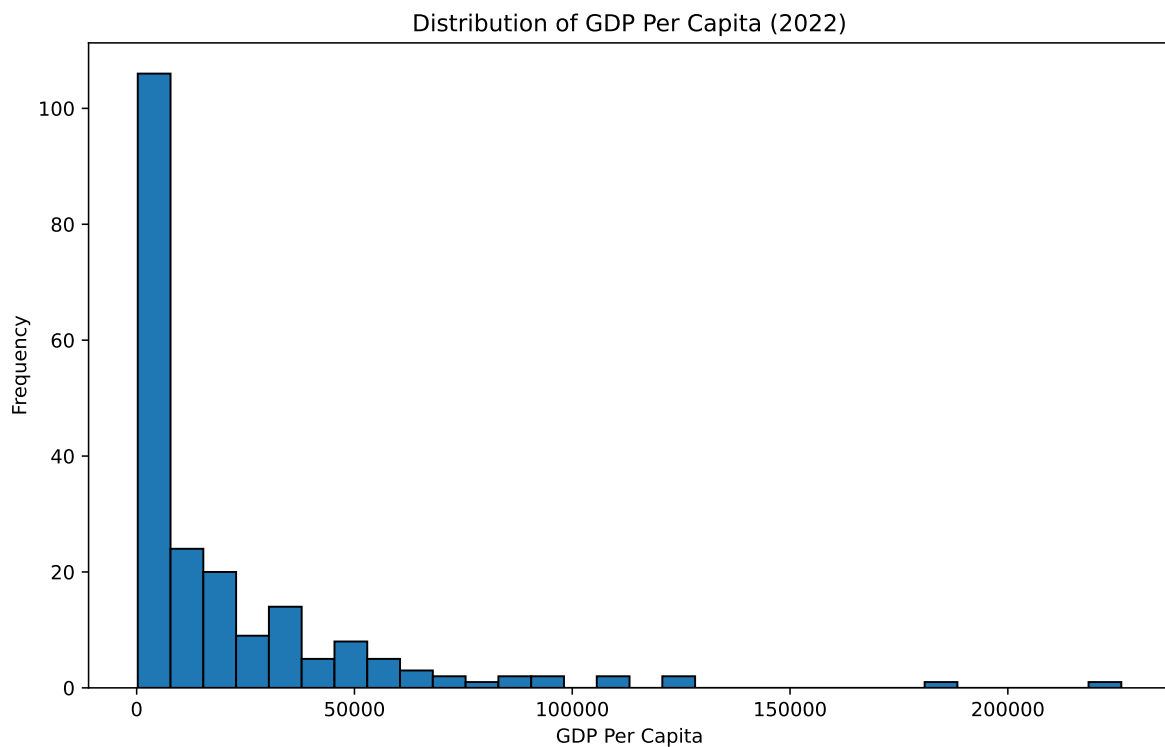
- **GDP per Capita:** The values vary significantly across countries, indicating large economic disparities.
- **Unemployment Rate:** The variation in unemployment rates across

countries may highlight differences in labor market conditions. - **Education Expenditure (as % of GDP):** This indicator shows how much countries invest in education relative to their GDP.

Plots

Distribution of GDP Per Capita

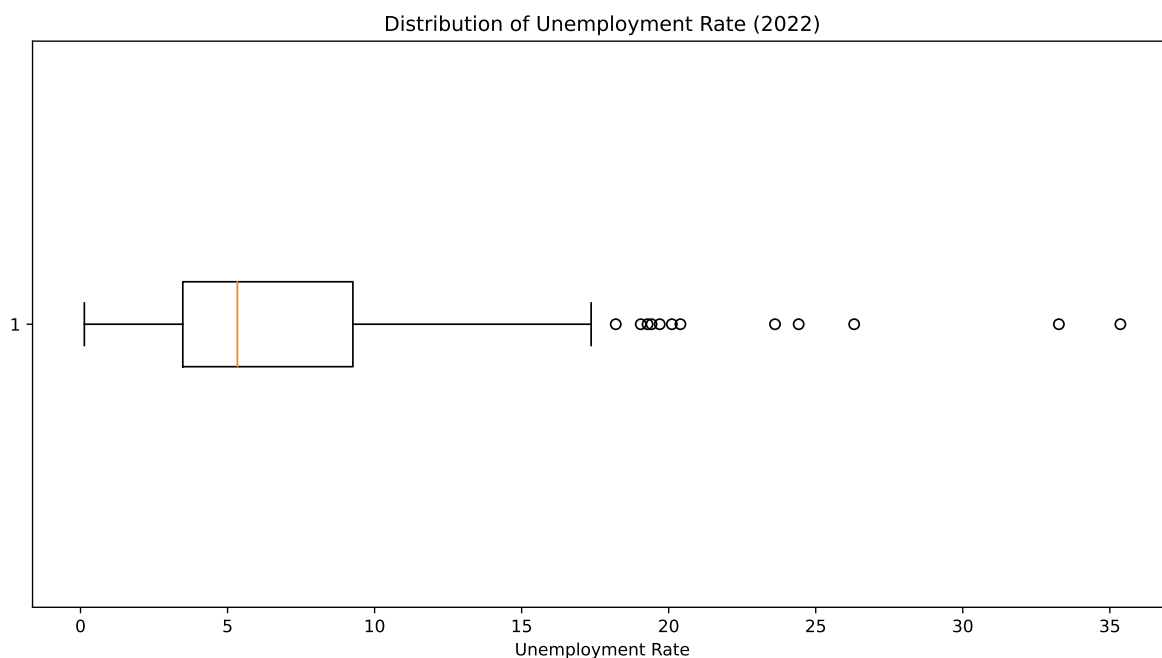
```
plt.figure(figsize=(10, 6))
data = df['gdp_per_capita'].dropna()
n, bins = np.histogram(data, bins=30)
plt.hist(bins[:-1], bins, weights=n, edgecolor='black')
plt.xlabel('GDP Per Capita')
plt.ylabel('Frequency')
plt.title('Distribution of GDP Per Capita (2022)')
plt.show()
```



The histogram (**Figure ?@fig-gdp**) shows that GDP per capita is right-skewed, meaning a few countries have exceptionally high GDP per capita, while most have lower values. The kernel density estimate (KDE) highlights the peak distribution. Economic inequality is evident.

Unemployment Rate Across Countries

```
plt.figure(figsize=(12, 6))
plt.boxplot(df['unemployment_rate'].dropna(), vert=False)
plt.xlabel('Unemployment Rate')
plt.title('Distribution of Unemployment Rate (2022)')
plt.show()
```

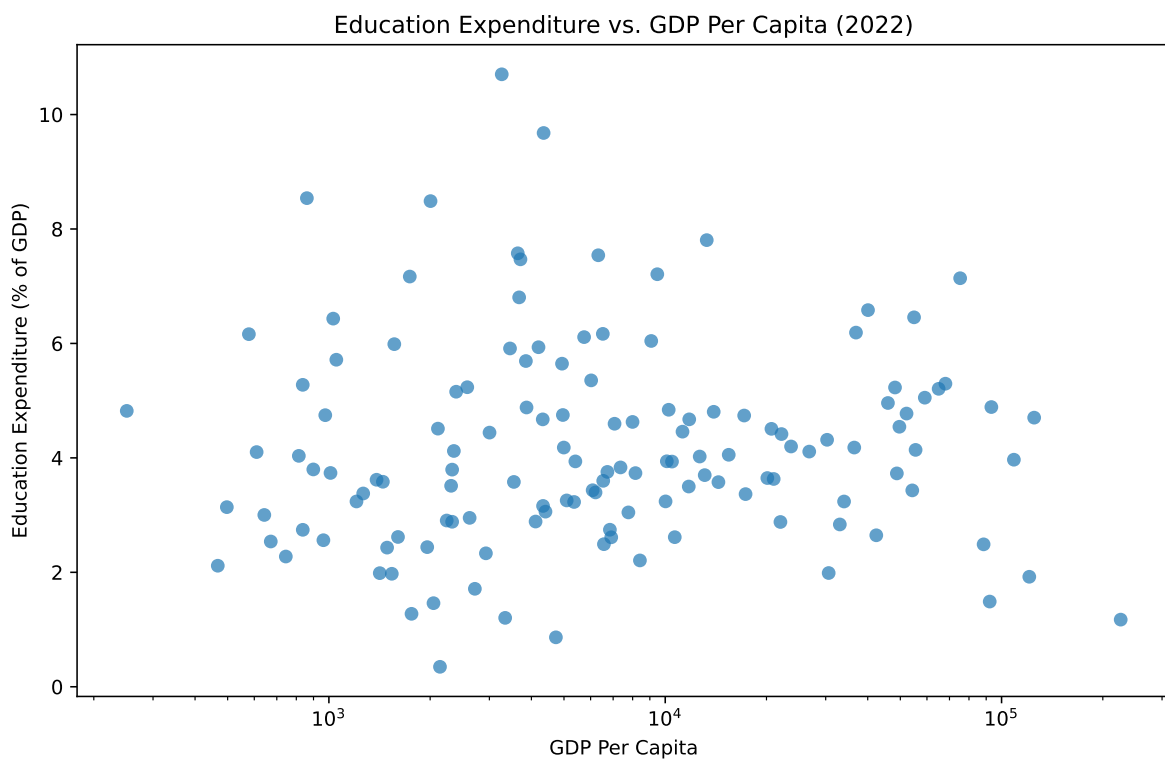


The box plot (**Figure ?@fig-unemployment**) shows the variation in unemployment rates among countries. Some nations experience very low unemployment, while others have significantly high rates, suggesting diverse economic conditions.

4. Education Expenditure vs. GDP Per Capita

```
plt.figure(figsize=(10, 6))
# Ensure both x and y values are the same size
gdp_per_capita = df_selected['gdp_per_capita']
edu_expenditure = df_selected['education_expenditure_gdp_share']

# Scatter plot
plt.scatter(gdp_per_capita, edu_expenditure, alpha=0.7)
plt.xlabel('GDP Per Capita')
plt.ylabel('Education Expenditure (% of GDP)')
plt.title('Education Expenditure vs. GDP Per Capita (2022)')
plt.xscale('log') # Log scale for better visualization
plt.show()
```



The scatter plot (**Figure ?@fig-education**) suggests whether wealthier countries invest a larger proportion of their GDP in education. A weak or strong correlation could reveal trends in educational funding relative to economic status.

Overall, this analysis highlights economic disparities, labor market conditions, and investment in education, providing insight into global development patterns.

References

```
@article{fuchs2010education,  
  title={Education or wealth: which matters more for reducing child mortality in developing countries},  
  author={Fuchs, Regina and Pamuk, Elsie and Lutz, Wolfgang},  
  journal={Vienna Yearbook of Population Research},  
  pages={175--199},  
  year={2010},  
  publisher={JSTOR}  
}  
  
@article{ginsburg2012public,  
  title={Public-private partnerships and the global reform of education in less wealthy countries},  
  author={Ginsburg, Mark},  
  journal={Comparative Education Review},  
  volume={56},  
  number={1},  
  pages={155--175},  
  year={2012},  
  publisher={University of Chicago Press Chicago, IL}  
}
```