Homework5

Rahul Ulman

2025-02-26

```
#Import the libraries
import pandas as pd
import wbgapi as wb
import matplotlib.pyplot as plt
import seaborn as sns
```

```
# 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=

# 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

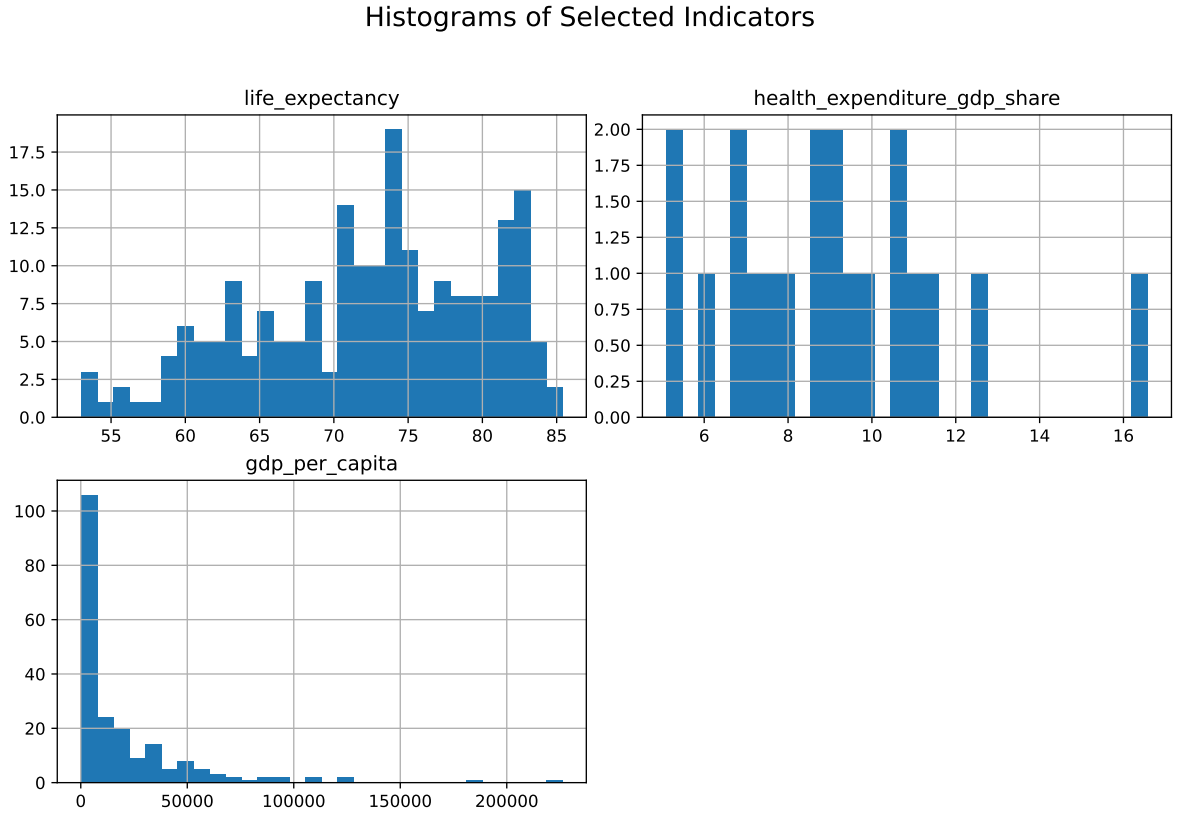

Figure 1: histogram

I picked these three indicators because I wanted to explore whether life expectancy of a country would be correlated to health expenditures, the GDP, niether or both. From this initial analysis, there is little to no correlation between life expectancy and either one of the other two variables, though there is slightly more positive correlation with health expenditure. This is visible in both (**heatmap?**) and (**pairplot?**). There is a positive correlation between GDP and health expenditure, which makes sense intuitively, as a country makes more money, more of that money can be turned towards keeping people healthy rather than investing in industry or other sectors.

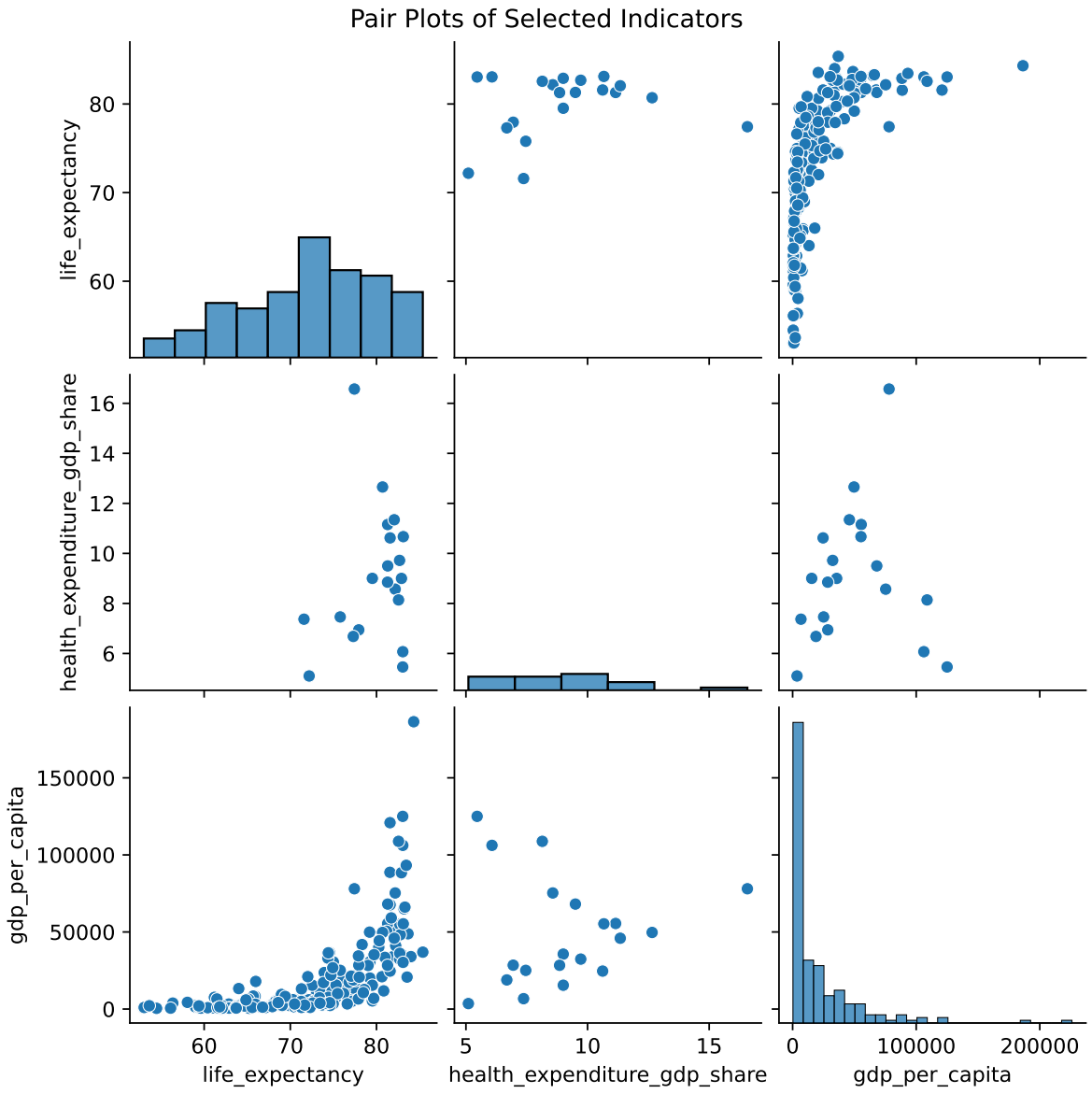

Figure 2: pairplot

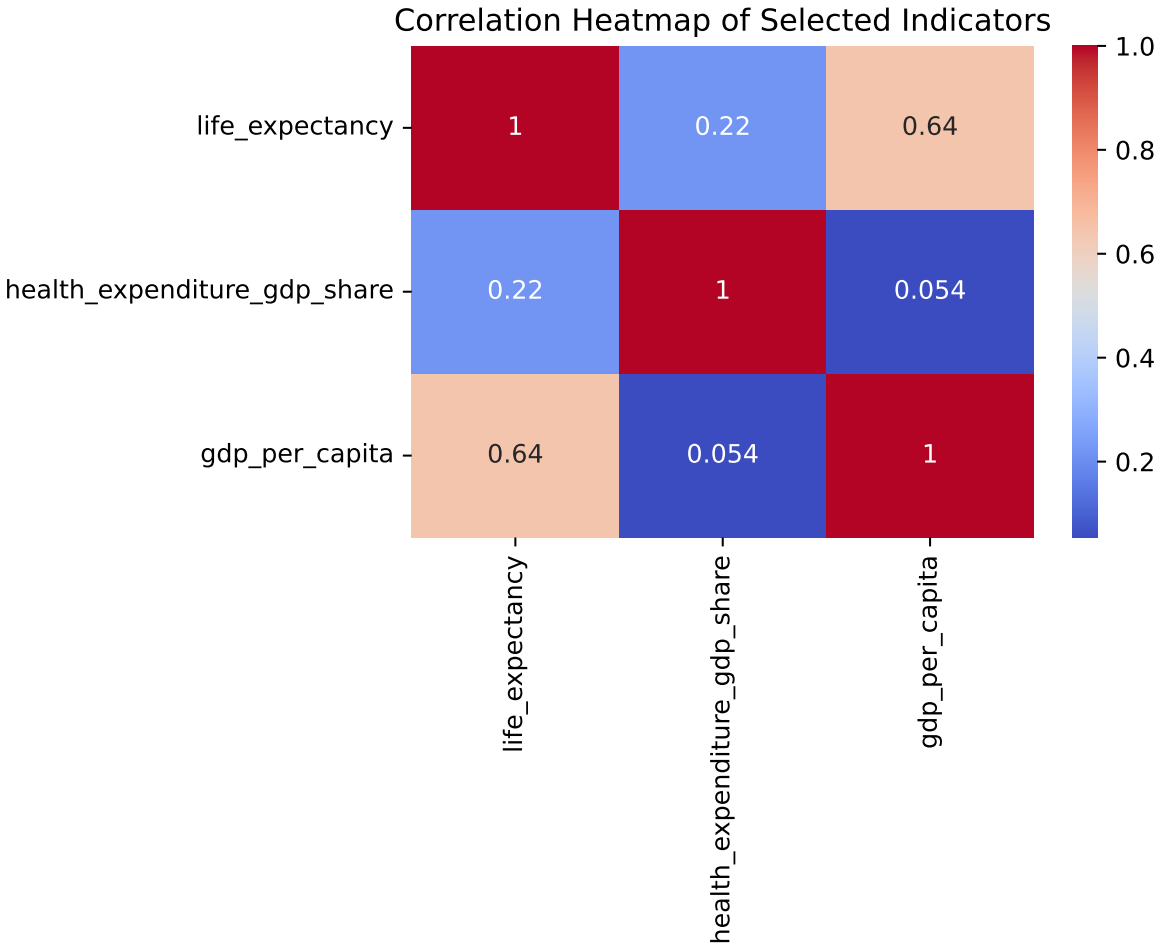

Figure 3: heatmap

```python
# Calculate summary statistics
summary_stats = df[['life_expectancy', 'health_expenditure_gdp_share', 'gdp_per_capita']].des

# Create a table with key statistics
key_stats = summary_stats.loc[['mean', 'std', 'min', '25%', '50%', '75%', 'max']]

# Display the table
print(key_stats)
```

```
      life_expectancy  health_expenditure_gdp_share  gdp_per_capita
mean        72.416519                      9.044045    20520.336828
std          7.713322                      2.703549    30640.741594
min         52.997000                      5.100000      250.634225
25%         66.782000                      7.263266     2599.752468
50%         73.514634                      8.925000     7606.237525
75%         78.475000                     10.632500    27542.145523
max         85.377000                     16.571152   226052.001905
```