Assignment 05 Report

Data

This report analyzes a sample of World Development Indicators (2022) (World Bank 2022).
The dataset includes 217 countries and 14 variables.

```python
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv("wdi.csv")
df.head()
```

	country	inflation_rate	exports_gdp_share	gdp_growth_rate	gdp_per_capita	adult_lite
0	Afghanistan	NaN	18.380042	-6.240172	352.603733	NaN
1	Albania	6.725203	37.395422	4.856402	6810.114041	98.5
2	Algeria	9.265516	31.446856	3.600000	5023.252932	NaN
3	American Samoa	NaN	46.957520	1.735016	19673.390102	NaN
4	Andorra	NaN	NaN	9.563798	42350.697069	NaN

Summary

```python
df[["gdp_per_capita", "life_expectancy", "income_inequality"]].describe()
```

	gdp_per_capita	life_expectancy	income_inequality
count	203.000000	209.000000	28.000000
mean	20345.707649	72.416519	38.328571
std	31308.942225	7.713322	7.721488
min	259.025031	52.997000	26.400000

	gdp_per_capita	life_expectancy	income_inequality
25%	2570.563284	66.782000	32.900000
50%	7587.588173	73.514634	38.100000
75%	25982.630050	78.475000	43.125000
max	240862.182448	85.377000	54.800000

The dataset includes 203 observations for GDP per capita, 209 for life expectancy, and only 28 for income inequality. This large difference in sample size suggests substantial missing data, especially for income inequality, which limits the reliability of comparisons involving that variable. The average GDP per capita is approximately $20,346, while the median is much lower at about $7,588, indicating a highly right-skewed distribution driven by a small number of very wealthy countries. Life expectancy is more tightly distributed, with a mean of about 72.4 years and relatively low variation. Income inequality has a mean of 38.3, with moderate spread across available countries.

```
count        203.000000
mean       20345.707649
std        31308.942225
min          259.025031
25%         2570.563284
50%         7587.588173
75%        25982.630050
max       240862.182448
Name: gdp_per_capita, dtype: float64
```

GDP per capita shows a very uneven distribution across countries. The mean value is approximately $20,346, but the median is only about $7,588, which suggests strong right skewness. The maximum value reaches $240,862, while the minimum is only $259, indicating extremely large disparities in income levels. The standard deviation (about $31,309) is even larger than the mean, further confirming substantial variation. The wide gap between the 25th percentile ($2,571) and the 75th percentile ($25,983) highlights global income inequality between lower-income and higher-income economies.

```
count     209.000000
mean       72.416519
std         7.713322
min        52.997000
25%        66.782000
50%        73.514634
75%        78.475000
```

```
max        85.377000
Name: life_expectancy, dtype: float64
```

Life expectancy is much more concentrated than GDP per capita. The mean is approximately
72.4 years and the median is about 73.5 years, suggesting a relatively symmetric distribution.
Values range from about 53.0 years to 85.4 years, showing noticeable but not extreme variation.
The standard deviation is about 7.7 years, indicating moderate dispersion across countries.
Compared to GDP, life expectancy appears less skewed and more clustered around the middle
range, reflecting global improvements in health outcomes.

```
count    28.000000
mean     38.328571
std       7.721488
min      26.400000
25%      32.900000
50%      38.100000
75%      43.125000
max      54.800000
Name: income_inequality, dtype: float64
```

Income inequality has only 28 non-missing observations, which significantly limits the scope
of analysis. The mean inequality index is approximately 38.3, with values ranging from 26.4
to 54.8. The distribution appears relatively centered, as the median (38.1) is very close to
the mean. The standard deviation of about 7.7 suggests moderate variation among countries
with available data. However, because of the small sample size, conclusions regarding global
inequality patterns should be interpreted cautiously.

Visualization

Table 3: Key statistics for selected indicators (2022). Source: World Bank WDI sample.

	mean	std	min	max
gdp_per_capita	20345.707649	31308.942225	259.025031	240862.182448
life_expectancy	72.416519	7.713322	52.997000	85.377000
income_inequality	38.328571	7.721488	26.400000	54.800000

The scatter plot in Figure 1 shows a clear positive relationship between GDP per capita and
life expectancy. As income increases, life expectancy generally rises, though the relationship
appears nonlinear and flattens at higher income levels. Countries with very low GDP per capita
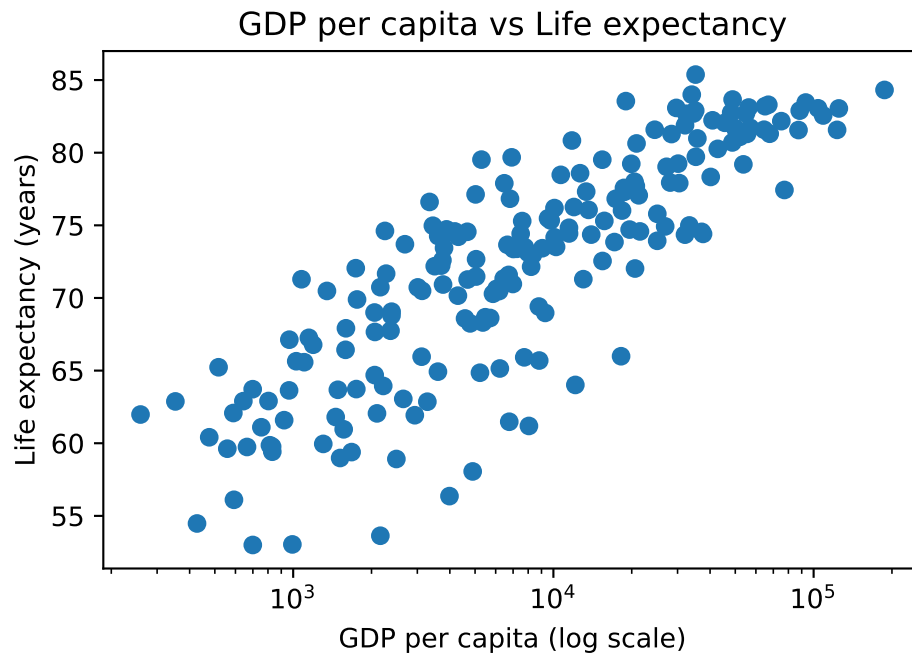tend to have substantially lower life expectancy, while wealthier countries cluster between

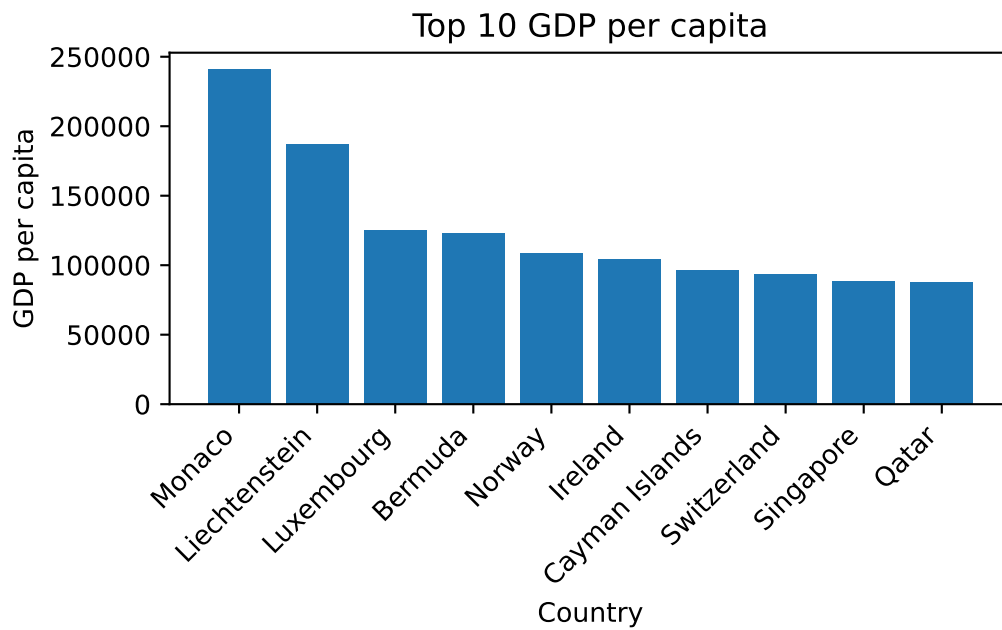Figure 1: GDP per capita vs life expectancy (2022). Source: World Bank WDI sample.

Figure 2: Top 10 countries by GDP per capita (2022). Source: World Bank WDI sample.

roughly 78 and 85 years. The use of a log scale makes this pattern more visible and highlights diminishing returns to income at higher levels.

Figure Figure 2 shows that a small group of countries have extremely high GDP per capita, with Monaco and Liechtenstein standing far above most others. This confirms the strong right skew observed in the summary statistics, where the mean GDP per capita is much higher than the median. However, the analysis has several limitations. Income inequality data are missing for most countries, reducing the reliability of comparisons involving that variable. Additionally, the scatter plot shows correlation rather than causation, and other factors such as healthcare systems, education, and public policy likely influence life expectancy beyond income alone. The relationship between income and life expectancy has been widely studied in demographic research (Preston 1975).

References

Preston, Samuel H. 1975. "The Changing Relation Between Mortality and Level of Economic Development." *Population Studies* 29 (2): 231–48.

World Bank. 2022. "World Development Indicators." https://databank.worldbank.org/source/world-development-indicators.