GLOBAL LIFE EXPECTANCY ANALYSIS

Presented By:

Dhruv Patel, Cole Price, Clayton Tucker, Caleb Damron, Danyil Chuprynov

Objective

- **Main Research Question:**
  What social, economic, and environmental factors were most associated with life expectancy across countries before the COVID-19 pandemic?
- **Key Goal:**
  Build a unified dataset and identify global patterns, correlations, and outliers.

DATA OVERVIEW

- **Sources Used (all pre-2020):**
  World Bank, WHO, UNICEF, World Population Review, Wikipedia, Kaggle.
- **Master Dataset:**
  193 countries, 9 factors (life expectancy, GDP, clean water access, health spending, happiness score, air quality, freedom, region, and income category)

	A	B	C	D	E	F	G	H	I	J
1	**Country**	**Life Expectancy**	**GDP**	**Clean Water Access %**	**Health Expenditure % GDP**	**Air Quality Score**	**Happiness Score (out of 10)**	**Freedom Score (out of 100)**	**Region**	**Income Category**
2	Afghanistan	62.94	18799444490.11	27.49	14.83	16.98	3.58	6	Middle East & North Africa	Low income
3	Albania	79.47	15585111614.04	70.67	6.86	12.60	4.96	68	Europe & Central Asia	Upper middle income
4	Algeria	75.68	193459662090.68	73.95	5.42	6.14	5.61	32	Middle East & North Africa	Upper middle income
5	Andorra	84.10	3155149347.81	90.64	7.32	8.72	6.52	93	Europe & Central Asia	High income
6	Angola	63.05	70897962732.03	55.50	2.48	21.07	4.03	28	Sub-Saharan Africa	Lower middle income
7	Antigua and Barbuda	77.17	1726448148.15	98.40	4.41	1.79	6.35	85	Latin America & Caribbean	High income
8	Argentina	76.85	447754683615.23	57.20	10.16	12.44	6.57	85	Latin America & Caribbean	Upper middle income
9	Armenia	76.22	13619290539.21	82.82	11.38	19.51	4.35	54	Europe & Central Asia	Upper middle income
10	Australia	82.90	1392723834562.64	96.50	10.21	3.36	7.28	95	East Asia & Pacific	High income
11	Austria	81.90	442983642371.84	98.90	10.49	9.52	7.20	93	Europe & Central Asia	High income
12	Azerbaijan	73.29	48174235294.12	70.16	4.18	11.87	5.21	7	Europe & Central Asia	Upper middle income
13	Bahrain	80.47	40446808510.64	98.91	3.96	22.75	5.96	12	Middle East & North Africa	High income
14	Bangladesh	72.63	351231654603.69	58.18	2.26	54.17	4.69	40	South Asia	Lower middle income
15	Barbados	76.50	5788288000.00	98.50	6.12	2.00	6.30	94	Latin America & Caribbean	High income

DATA CLEANING & INTEGRATION

- 8 datasets merged
- Standardized country names (e.g., *"Republic of Korea" → "South Korea"*)
- Filled missing values using alternative reputable sources
- Unified into a single master table

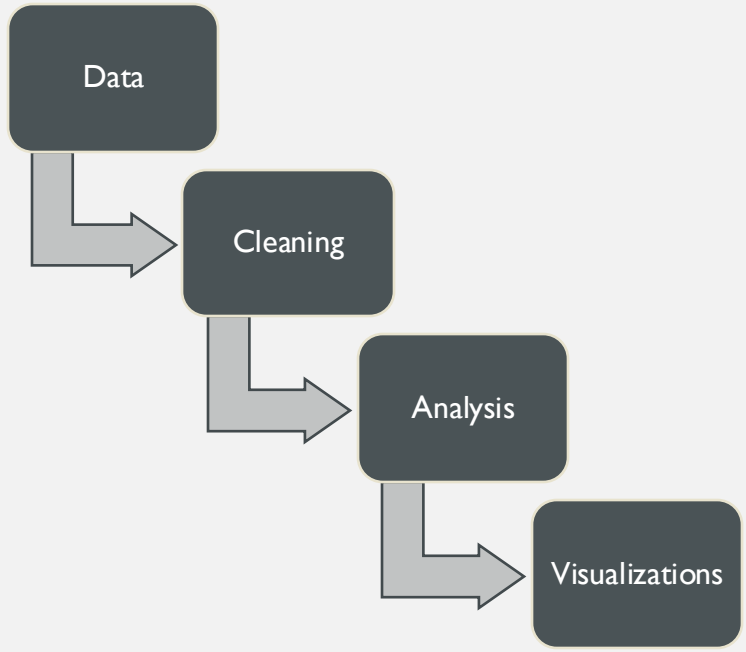
```python
import pandas as pd

life_raw = pd.read_csv("Life Expectancy (World Bank).csv", skiprows=4)
gdp_raw = pd.read_csv("GDP (World Bank).csv", skiprows=4)
water_raw = pd.read_csv("Safe Drinking Water Access (World Bank).csv", skiprows=4)
health_raw = pd.read_csv("Health Expenditure % GDP (World Bank).csv", skiprows=4)
happiness_raw = pd.read_csv("Happiness Index (Kaggle).csv")
freedom_raw = pd.read_csv("Freedom Index (World Population Review).csv")
air_raw = pd.read_csv("Air Quality (Wikipedia).csv")
income_raw = pd.read_csv("Income Group and Region (World Bank).csv")

country_fix = {
    "Republic of Korea": "South Korea",
    "Korea, Rep.": "South Korea",
    "Korea, South": "South Korea",
    "Korea, Democratic People's Rep.": "North Korea",
    "Russian Federation": "Russia",
    "Egypt, Arab Rep.": "Egypt",
    "Gambia, The": "Gambia",
    "Iran, Islamic Rep.": "Iran",
    "Venezuela, RB": "Venezuela",
    "Congo, Dem. Rep.": "Democratic Republic of the Congo",
    "Congo, Rep.": "Republic of the Congo",
    "Bahamas, The": "Bahamas",
    "Yemen, Rep.": "Yemen"
}

def map_country(series):
    return series.apply(lambda x: country_fix.get(x, x))
```

METHODS

We used:

- Descriptive statistics
- Pearson correlations
- Grouping by region & income
- Outlier detection

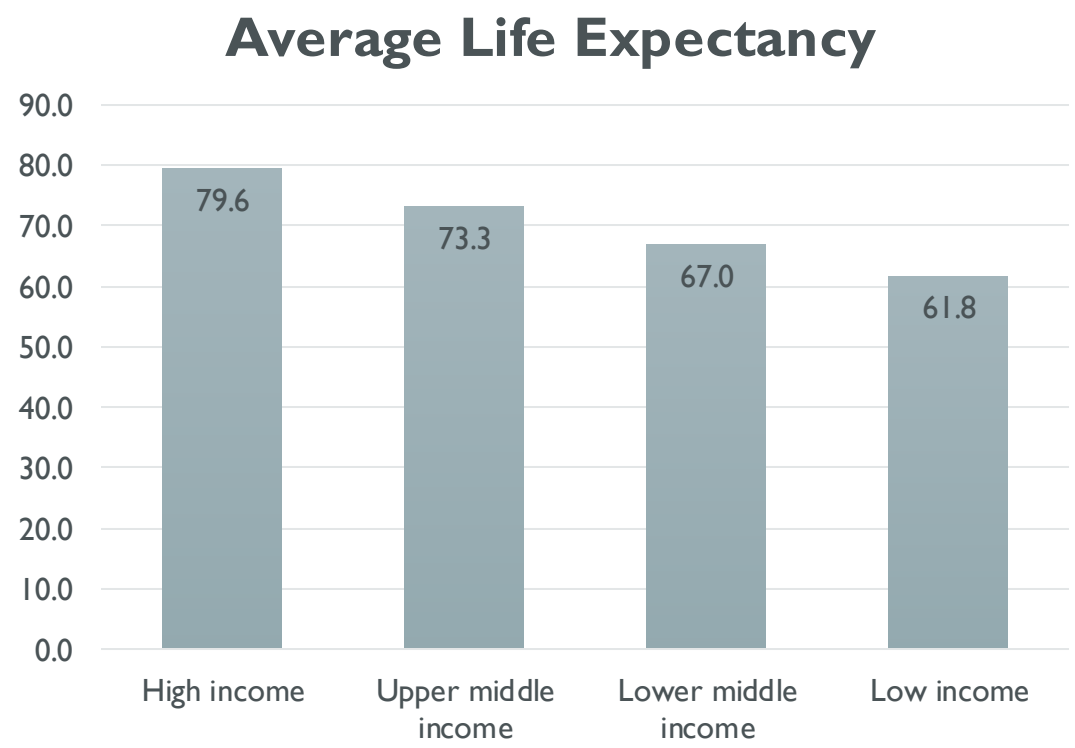
Goal: reveal broad global patterns, not build predictions.



KEY RESULTS: GLOBAL TRENDS

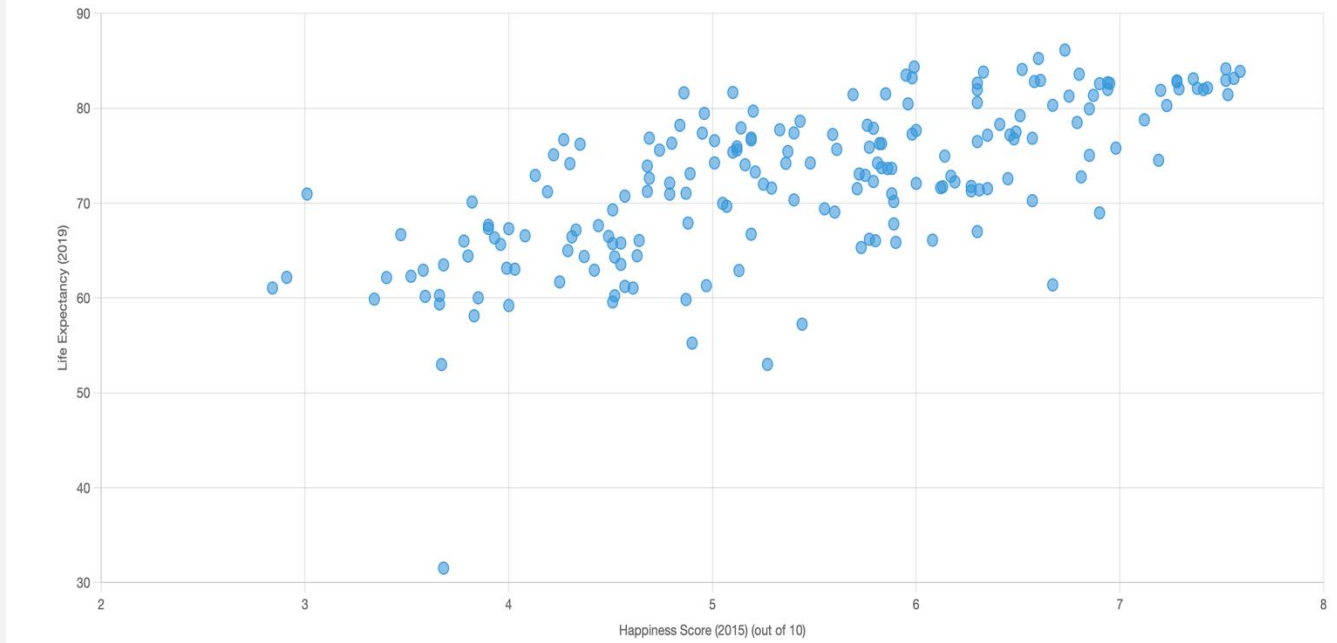
Average life expectancy across all countries: **72.4 years**

- High-income ≈ 79.6
- Upper-middle ≈ 73.3
- Lower-middle ≈ 67.0
- Low-income ≈ 61.8



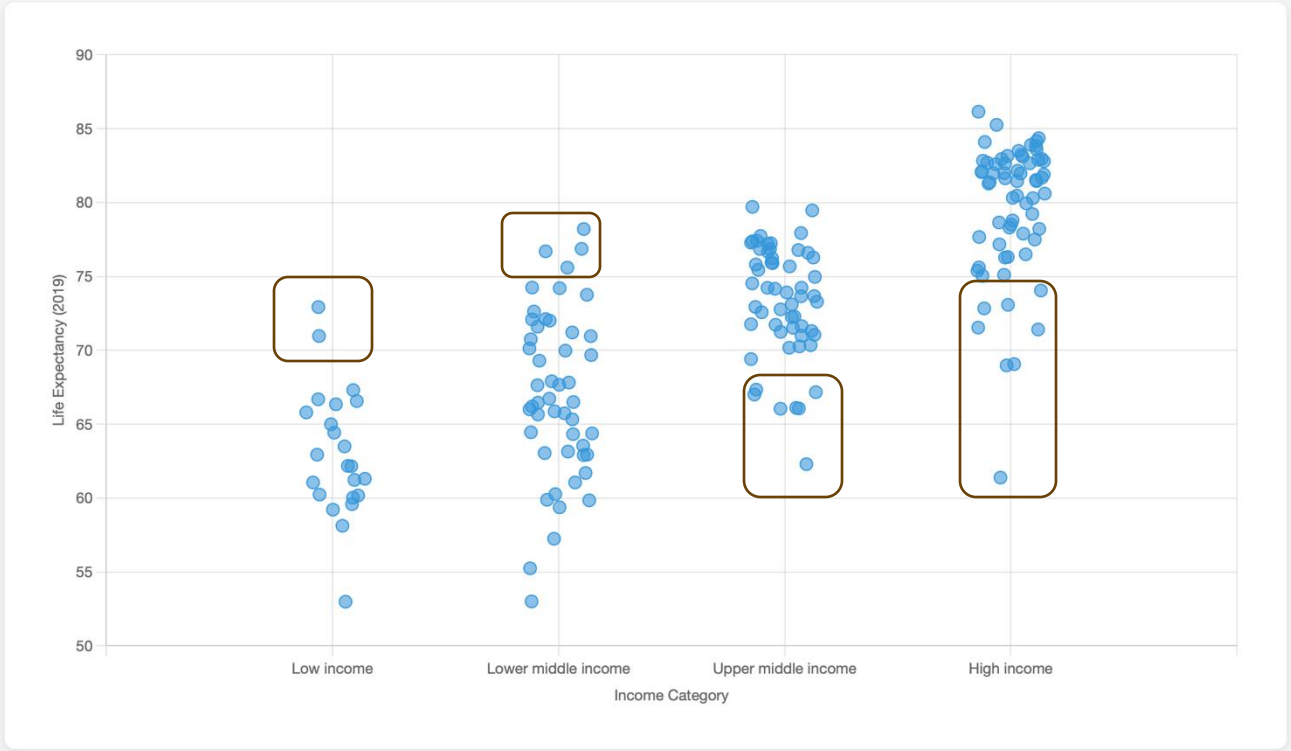
CORRELATIONS

- Highest correlations with life expectancy:
  - +0.69 Clean water access
  - +0.69 Happiness
  - +0.48 Freedom
- Weaker correlations:
  - +0.20 Health expenditure
  - +0.18 GDP
- Negative correlation:
  - –0.20 Air pollution *(Lower = better)*



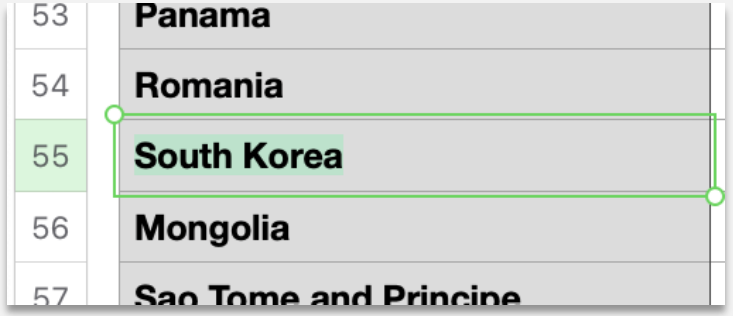
OUTLIERS

- High-performing low income:
  - Syria, North Korea
- High-performing lower-middle income:
  - Jordan, Lebanon, Sri Lanka, Tunisia
- Low-performing upper-middle income:
  - Fiji, Marshall Islands, South Africa, Tuluva, Botswana, Guinea, Gabon
- Low-performing high-income:
  - Nauru, Guyana, Bahamas, Trinidad & Tobago, Russia, Seychelles



ISSUES ENCOUNTERED

- Inconsistent naming conventions
- Missing factors across datasets
- Different years (2015–2019)
- Manual validation / missing data filling required



FUTURE WORK

01

Add GDP per capita + inequality metrics

02

Use time-series & post-COVID data

03

Add environmental + health indicators (PM2.5, smoking, chronic disease)

04

Build predictive models

05

Expand the interactive dashboard

THANK YOU!

Questions?

Presented By:

Dhruv Patel, Cole Price, Clayton Tucker, Caleb Damron, Danyil Chuprynov