Does urbanization drive changes in poverty in Afghanistan (2006–2022)?

Question

Does urbanization drive changes in poverty in Afghanistan from 2006 to 2022?
We use **slum population (% of urban population)** as a proxy for urbanization and **poverty headcount ratio (% of population)** to measure poverty.

Data

- Source: World Bank WDI
- Years: 2006–2022
- Indicators:
  - Population living in slums (% of urban population)
  - Poverty headcount ratio at national poverty lines (% of population)
- Notes: Only years with available data for each indicator are included.

Method

1. Filtered dataset for Afghanistan.
2. Selected relevant columns: Year, Indicator Name, Value.
3. Pivoted indicators into separate columns and sorted by year.
4. Produced a **line graph** showing slum population vs poverty headcount over time.

Results

- **Table:** Pivoted dataset (`data_clean/afghanistan_slums_vs_poverty.csv`)
- **Graph:** Line chart (`figures/afghanistan_slums_vs_poverty_lines.png`)

Observations:

- Slum population increased steadily from 2006–2022.
- Poverty headcount showed fluctuations over time.
- Visually, there is a possible correspondence between urbanization (slum population) and poverty changes, though causality cannot be inferred.

Interpretation

- Rising slum population may reflect urbanization pressures in Afghanistan.

- Trends suggest urbanization may be associated with poverty changes.
- The analysis is descriptive and does not establish causal relationships.

Limitations

- Only two indicators are analyzed.
- Data gaps exist for some years.
- National-level data only; no regional breakdown.
- No formal causal methods applied.

Next Steps / Extensions

- Analyze regional data if available.
- Include other urbanization or poverty indicators.
- Extend analysis to other countries for comparison.

In [20]:

```
# ============================
# POVERTY INDICATORS COMPARISON: AFGHANISTAN (LINES CONNECTED, LIMITED TO DA
# ============================

import pandas as pd
import matplotlib.pyplot as plt
import os

# Folders
data_raw_folder = "data_raw/"
data_clean_folder = "data_clean/"
figures_folder = "figures/"

# Load CSV
filename = "afghanistan_poverty.csv"
df = pd.read_csv(os.path.join(data_raw_folder, filename))

# Keep only Afghanistan + needed columns
df = df[df["Country Name"] == "Afghanistan"]
df = df[["Year", "Indicator Name", "Value"]]

# Convert Year and Value to numeric, drop invalid Value rows
df["Year"] = pd.to_numeric(df["Year"], errors="coerce")
df["Value"] = pd.to_numeric(df["Value"], errors="coerce")
df = df.dropna(subset=["Value"])

# Aggregate duplicates
df = df.groupby(["Year", "Indicator Name"], as_index=False).mean()

# Pivot indicators into separate columns
df_pivot = df.pivot(index="Year", columns="Indicator Name", values="Value").
df_pivot = df_pivot.sort_values("Year")

print("Pivoted dataset:")
```

```
display(df_pivot)

# Interpolate for plotting
df_plot = df_pivot.interpolate(method='linear')

# Identify the actual data ranges for each indicator
slum_years = df_pivot.dropna(subset=["Population living in slums (% of urbar
poverty_years = df_pivot.dropna(subset=["Poverty headcount ratio at national

# Plot two lines
plt.figure(figsize=(10,6))

# Slum population line (full range)
plt.plot(df_plot["Year"],
         df_plot["Population living in slums (% of urban population)"],
         marker='o', label="Slum Population (% of urban population)")

# Poverty headcount line (limited to actual data range)
poverty_mask = (df_plot["Year"] >= poverty_years.min()) & (df_plot["Year"] <
plt.plot(df_plot["Year"][poverty_mask],
         df_plot["Poverty headcount ratio at national poverty lines (% of po
         marker='o', label="Poverty Headcount (% of population)")

plt.title("Afghanistan: Slum Population vs Poverty Headcount")
plt.xlabel("Year")
plt.ylabel("Percentage of Population")
plt.xticks([2006,2008,2010,2012,2014,2016,2018,2020,2022])
plt.legend()
plt.grid(True)
plt.tight_layout()
plt.savefig(os.path.join(figures_folder, "afghanistan_slums_vs_poverty_lines
plt.show()

# Save cleaned CSV (original data, no interpolation)
df_pivot.to_csv(os.path.join(data_clean_folder, "afghanistan_slums_vs_povert
```

```
Pivoted dataset:
```

Indicator Name	Year	Multidimensional poverty headcount ratio (% of total population)	Multidimensional poverty headcount ratio, children (% of population ages 0-17)	Multidimensional poverty index (scale 0-1)	Multidimensional poverty index, children (population ages 0-17) (scale 0-1)
0	2006	NaN	NaN	NaN	NaN
1	2007	NaN	NaN	NaN	NaN
2	2008	NaN	NaN	NaN	NaN
3	2010	NaN	NaN	NaN	NaN
4	2011	NaN	NaN	NaN	NaN
5	2012	NaN	NaN	NaN	NaN
6	2014	NaN	NaN	NaN	NaN
7	2016	51.7	56.4	0.270	0.30
8	2018	NaN	NaN	NaN	NaN
9	2020	49.4	53.8	0.265	0.29
10	2022	NaN	NaN	NaN	NaN