Do countries with a declining birth rate also have a trend to urbanization?

Maximilian Eineder

Dataset

The World Development Indicators Dataset was used.

Motivation

I wanted to examine whether countries with a declining birth rate tend to have a trend to urbanization, meaning that the lower the birth rate in a country, the higher the degree of urban population. This might be interesting because if a country’s birth rate is declining, this could be a sign that migration into cities might take place in the future. With that knowledge, city planners could react appropriately.

Research Question

I wanted to examine whether countries with a declining birth rate tend to have a trend to urbanization.

Findings I

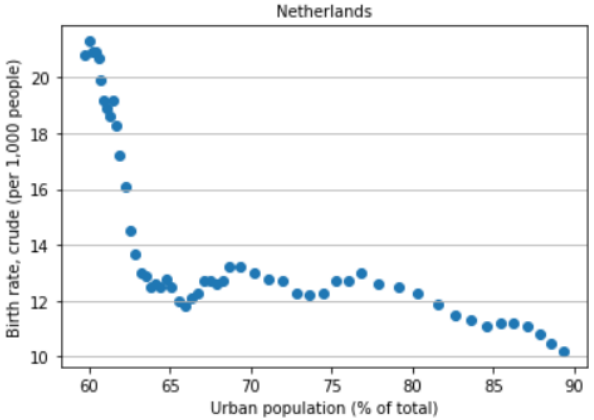
In general, the data supports the hypothesis that countries with a declining birth rate tend to have a trend to urbanization.

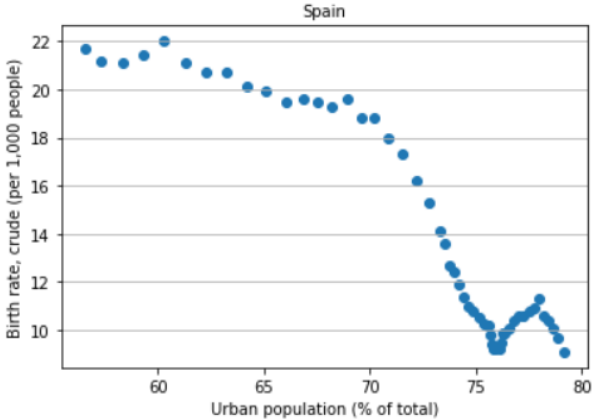
The correlation of the two indicators over all countries and years is considerably negative:

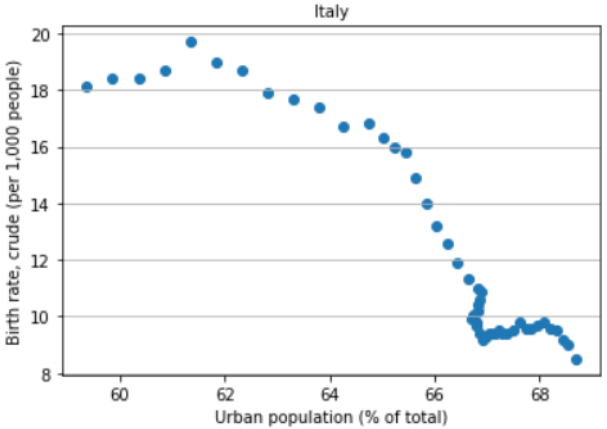
IndicatorName	Birth rate, crude (per 1,000 people)	Urban population (% of total)
IndicatorName		
Birth rate, crude (per 1,000 people)	1.000000	-0.682624
Urban population (% of total)	-0.682624	1.000000

Findings II

As typical examples, in European countries this negative correlation is clearly visible:

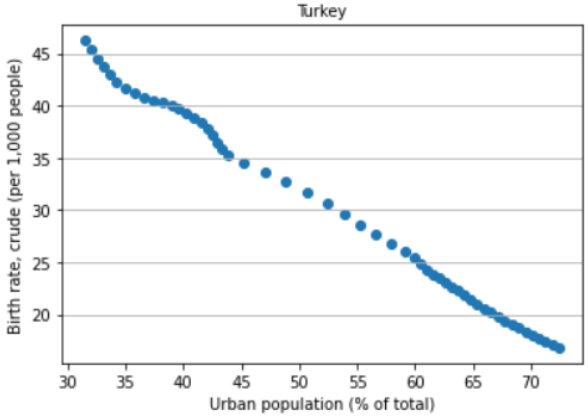


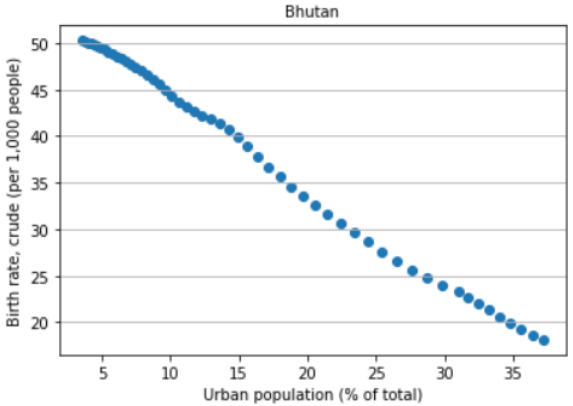


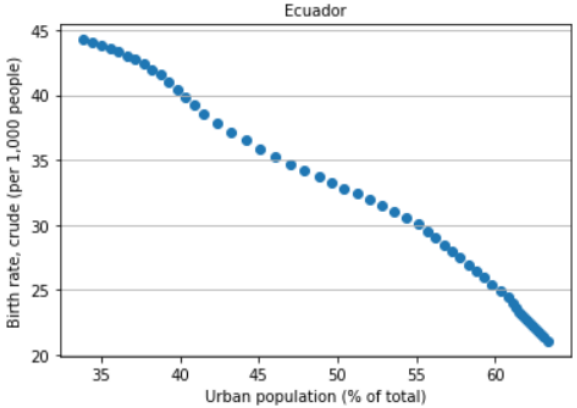


Findings III

These are the countries for which this correlation is lowest:

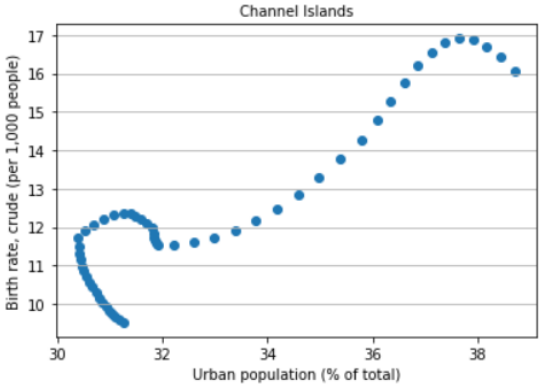


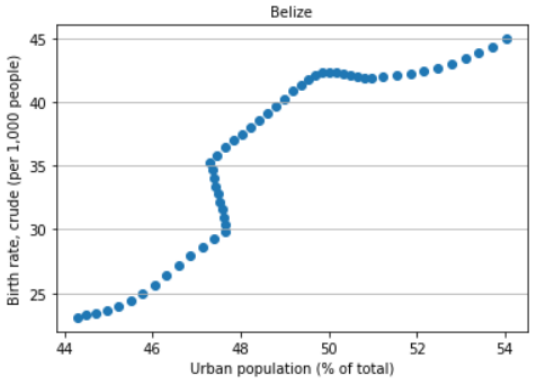


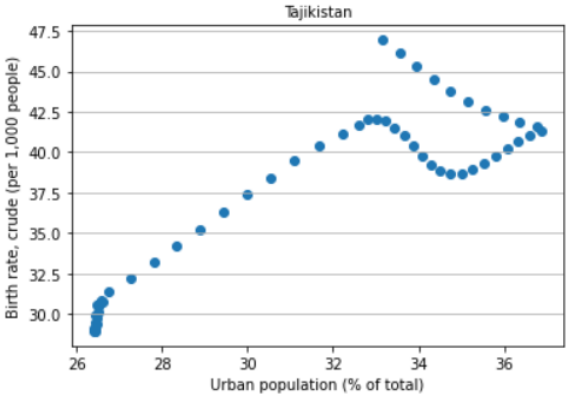


Findings IV

But there are also countries for which this correlation is positive:







Findings V

Conclusion: In general, these findings support the hypothesis that countries with a declining birth rate tend to have a trend to urbanization.

Acknowledgements

I would like to thank my girlfriend Marianne for the feedback on my work.

Let's see whether the two indicators

- Urban population (% of total)
- Birth rate, crude (per 1,000 people

correlate in general and if they correlate very well in specific countries.

In [86]:

```
import pandas as pd
import numpy as np
import random
import matplotlib.pyplot as plt
data = pd.read_csv('./world-development-indicators/Indicators.csv')
```

In [87]:

```
indicator1 = "Urban population (% of total)"
indicator2 = "Birth rate, crude (per 1,000 people)"
filter1 = data["IndicatorName"] == indicator1
filter2 = data["IndicatorName"] == indicator2
reduced_data = data[filter1 | filter2]
reduced_data = reduced_data.dropna()
pivot = pd.pivot_table(reduced_data, index=['Year', 'CountryName'], columns='IndicatorName', values='Value', aggfunc='mean')
pivot.corr()
```

Out[87]:

IndicatorName	Birth rate, crude (per 1,000 people)	Urban population (% of total)
IndicatorName		
Birth rate, crude (per 1,000 people)	1.000000	-0.682624
Urban population (% of total)	-0.682624	1.000000

Seems like the two indicators have a fairly strong negative correlation over all countries.

Let's find countries where the two indicators correlate even more.

In [88]:

```
pivot.head()
```

Out[88]:

	IndicatorName	Birth rate, crude (per 1,000 people)	Urban population (% of total)
Year	CountryName		
1960	Afghanistan	51.276	8.221
	Albania	42.240	30.705
	Algeria	49.671	30.510
	American Samoa	NaN	66.211
	Andorra	NaN	58.450

In [89]:

```
kosovo_filter = data["CountryName"] == "Kosovo"
data[kosovo_filter & filter2]
```

Out[89]:

	CountryName	CountryCode	IndicatorName	IndicatorCode	Year	Value
1140554	Kosovo	KSV	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	1981	31.0
1225274	Kosovo	KSV	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	1982	33.4
1310564	Kosovo	KSV	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	1983	30.8
1396148	Kosovo	KSV	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	1984	33.5
1482766	Kosovo	KSV	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	1985	32.1
1570520	Kosovo	KSV	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	1986	31.8
1659323	Kosovo	KSV	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	1987	32.1
1748508	Kosovo	KSV	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	1988	31.4
1839800	Kosovo	KSV	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	1989	29.4
1945826	Kosovo	KSV	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	1990	29.6
2062666	Kosovo	KSV	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	1991	27.5
2181233	Kosovo	KSV	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	1992	23.0
2304309	Kosovo	KSV	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	1993	22.5
2429121	Kosovo	KSV	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	1994	21.8
2559488	Kosovo	KSV	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	1995	22.1
2694606	Kosovo	KSV	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	1996	22.4
2830019	Kosovo	KSV	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	1997	20.6
2966922	Kosovo	KSV	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	1998	21.2
3106728	Kosovo	KSV	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	1999	22.7
3255913	Kosovo	KSV	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	2000	22.8
3407798	Kosovo	KSV	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	2001	21.7
3560817	Kosovo	KSV	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	2002	20.8
3715012	Kosovo	KSV	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	2003	18.3
3870728	Kosovo	KSV	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	2004	20.0

	CountryName	CountryCode	IndicatorName	IndicatorCode	Year	Value
4041472	Kosovo	KSV	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	2005	20.9
4219706	Kosovo	KSV	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	2006	19.1
4398761	Kosovo	KSV	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	2007	19.3
4578976	Kosovo	KSV	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	2008	19.0
4757883	Kosovo	KSV	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	2009	18.8
4941325	Kosovo	KSV	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	2010	18.5
5122264	Kosovo	KSV	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	2011	18.3
5297813	Kosovo	KSV	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	2012	18.0
5462688	Kosovo	KSV	Birth rate, crude (per 1,000 people)	SP.DYN.CBRT.IN	2013	17.7

In [90]:

```
countries = data["CountryName"].unique()
def get_corr(country):
    country_data = reduced_data[reduced_data["CountryName"] == country]
    country_piv = country_data.pivot(index='Year', columns='IndicatorName', values='Val
ue')
    country_piv = country_piv.dropna()
    if country_piv.shape[1] < 2:
        return None
    corr = country_piv.corr()[indicator1][indicator2]
    return corr
```

In [91]:

```
correlation_data = pd.DataFrame(columns=['Country', 'Correlation'])
i = 0
for country in countries:
    correlation = get_corr(country)
    correlation_data.loc[i] = [country] + [correlation]
    #print(country + ": " + str(correlation))
    i = i + 1

highest_correlation = correlation_data.sort_values(by=['Correlation']).head(5)
lowest_correlation = correlation_data.sort_values(by=['Correlation'], ascending=False).
head(5)
```

In [92]:

```
highest_correlation
```

Out[92]:

	Country	Correlation
228	Turkey	-0.998797
55	Bhutan	-0.998288
15	Latin America & Caribbean (developing only)	-0.996132
89	Ecuador	-0.995686
14	Latin America & Caribbean (all income levels)	-0.994660

In [93]:

```
lowest_correlation
```

Out[93]:

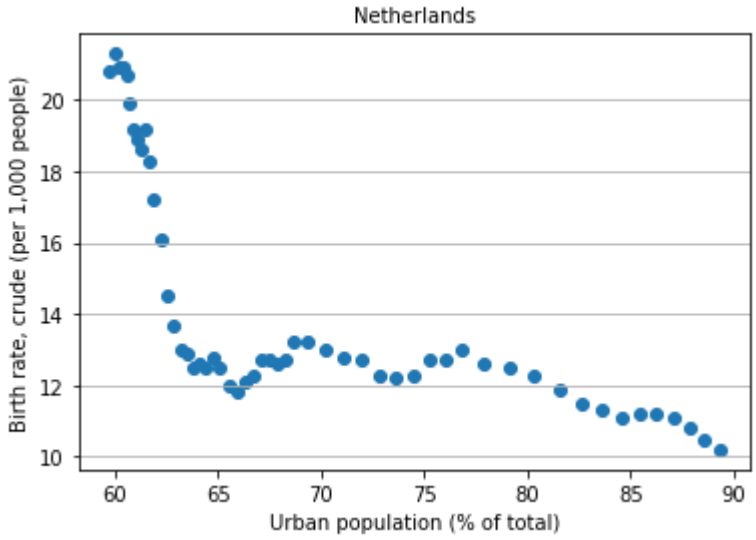
	Country	Correlation
36	American Samoa	1.000000
71	Channel Islands	0.923783
52	Belize	0.909631
220	Tajikistan	0.888484
144	Liechtenstein	0.875702

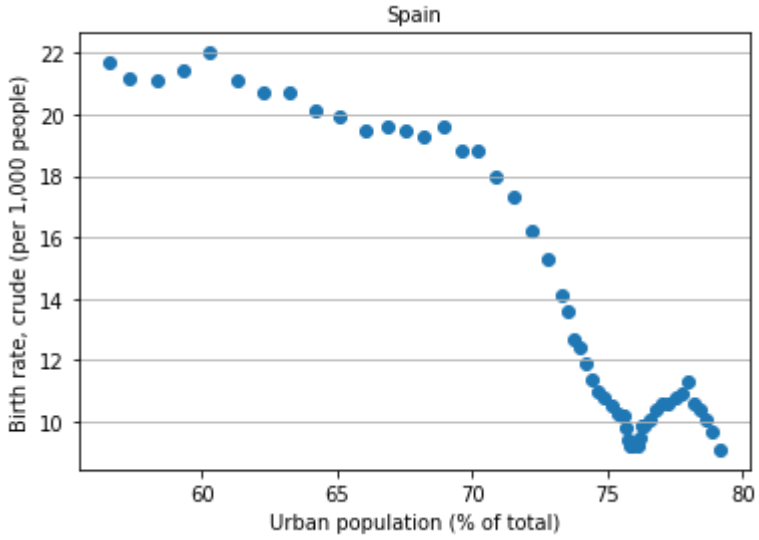
In [94]:

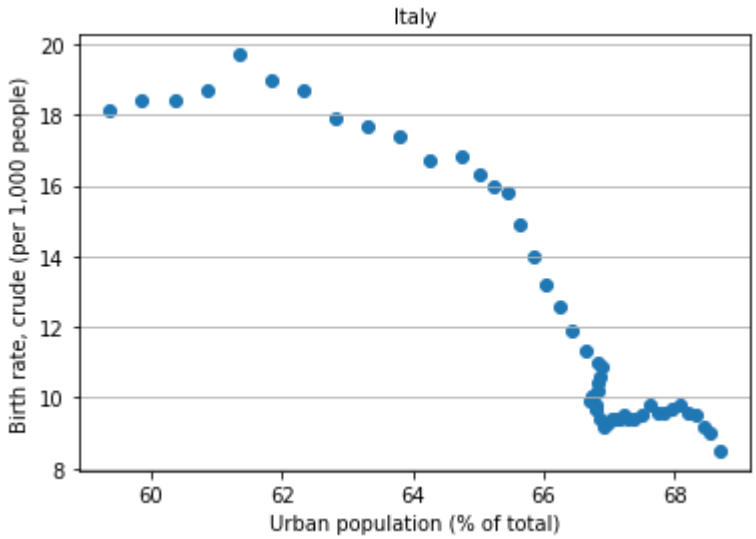
```
%matplotlib inline
import matplotlib.pyplot as plt

def plot_country(country_data, country):
    fig, axis = plt.subplots()
    # Grid lines, Xticks, Xlabel, Ylabel

    axis.yaxis.grid(True)
    axis.set_title(country,fontsize=10)
    axis.set_xlabel(indicator1,fontsize=10)
    axis.set_ylabel(indicator2,fontsize=10)

    X = country_data[indicator1]
    Y = country_data[indicator2]

    axis.scatter(X, Y)
```

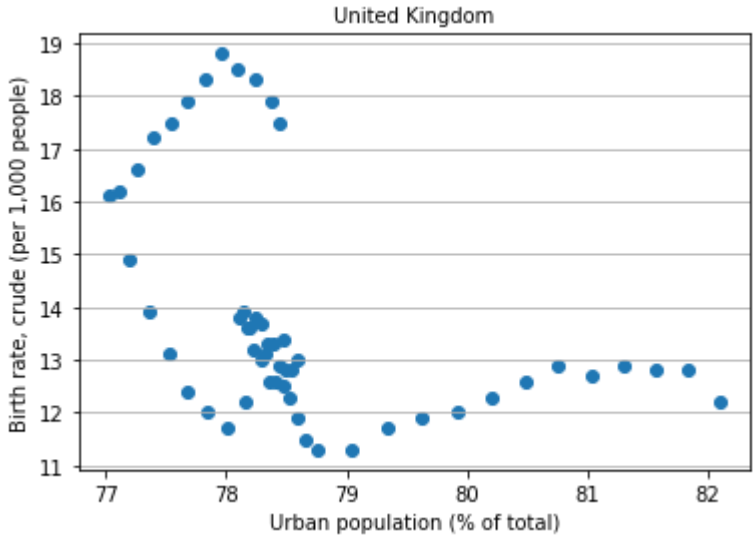
In [95]:

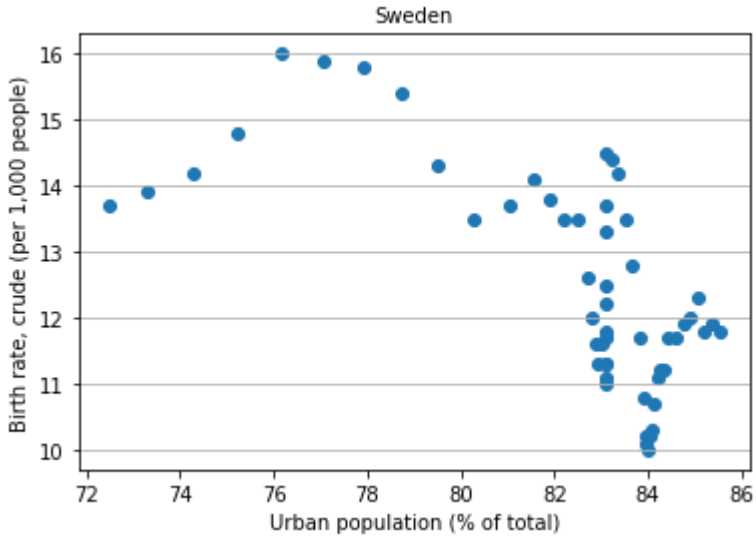
```
european_countries = ["Netherlands", "Spain", "Italy", "United Kingdom", "Sweden"]
for country in european_countries:
    country_data = reduced_data[reduced_data["CountryName"] == country]
    country_piv = country_data.pivot(index='Year', columns='IndicatorName', values='Val
ue')
    plot_country(country_piv, country)
```


Netherlands
Birth rate, crude (per 1,000 people)
Urban population (% of total)


Spain
Birth rate, crude (per 1,000 people)
Urban population (% of total)

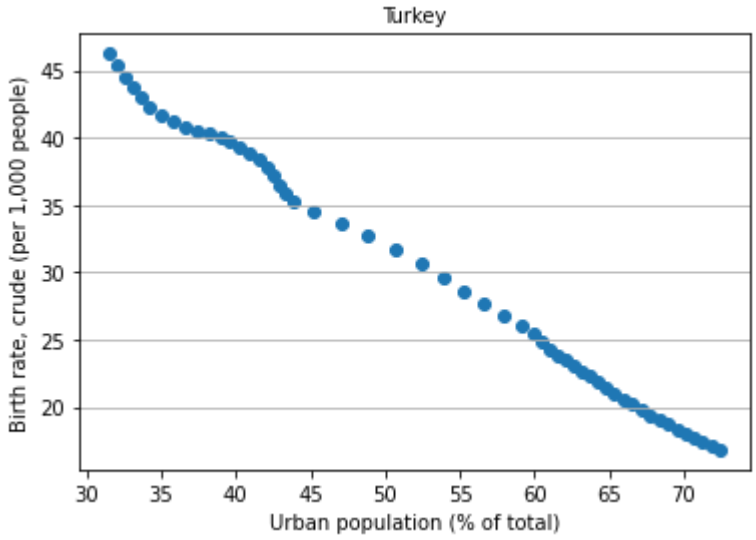

Italy
Birth rate, crude (per 1,000 people)
Urban population (% of total)

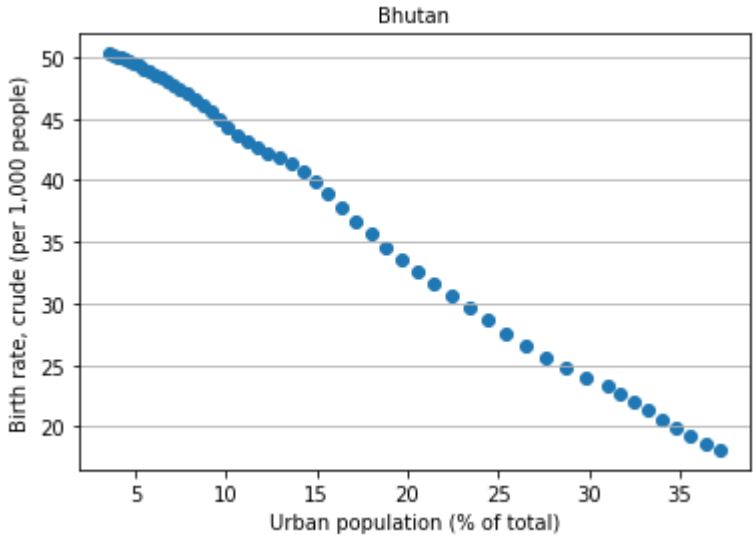

United Kingdom
Birth rate, crude (per 1,000 people)
Urban population (% of total)
19
18
17
16
15
14
13
12
11
77
78
79
80
81
82

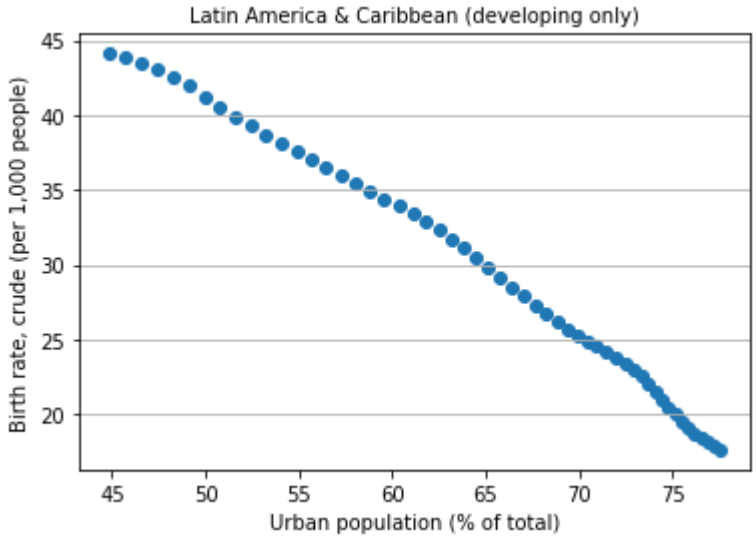

Sweden
Birth rate, crude (per 1,000 people)
Urban population (% of total)
16
15
14
13
12
11
10
72
74
76
78
80
82
84
86

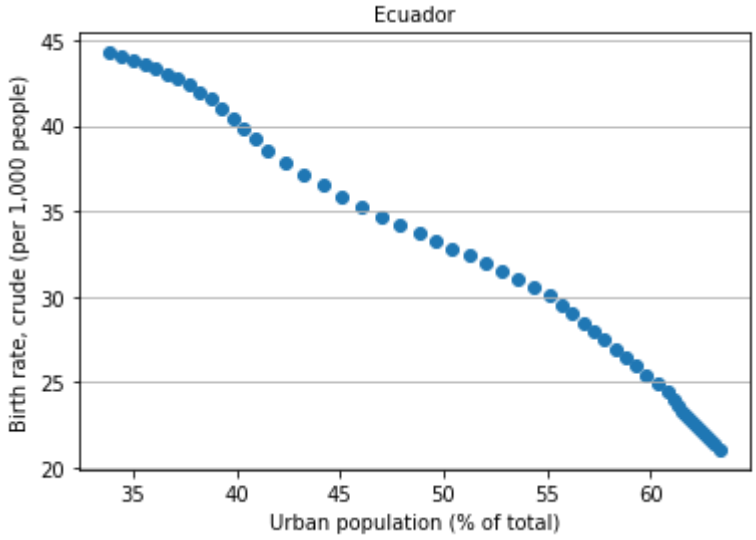
In [96]:

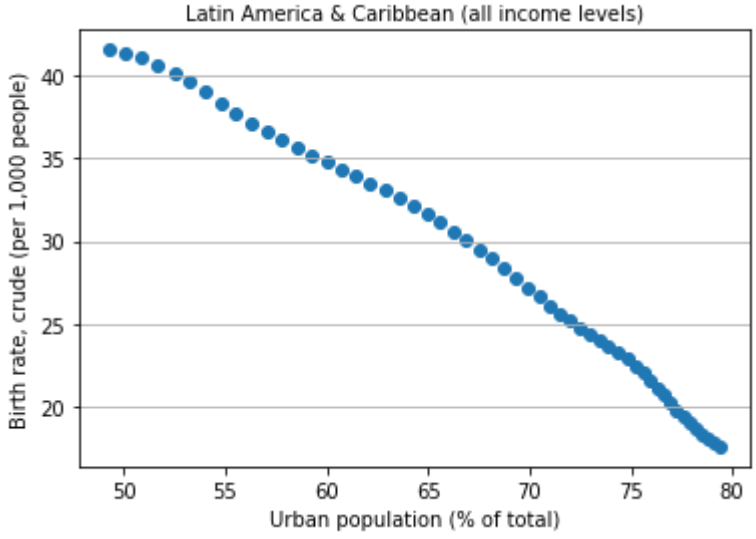
```
for country in highest_correlation["Country"]:
    country_data = reduced_data[reduced_data["CountryName"] == country]
    country_piv = country_data.pivot(index='Year', columns='IndicatorName', values='Val
ue')
    plot_country(country_piv, country)
```


Turkey
Birth rate, crude (per 1,000 people)
45
40
35
30
25
20
30
35
40
45
50
55
60
65
70
Urban population (% of total)


Bhutan
Birth rate, crude (per 1,000 people)
50
45
40
35
30
25
20
5
10
15
20
25
30
35
Urban population (% of total)

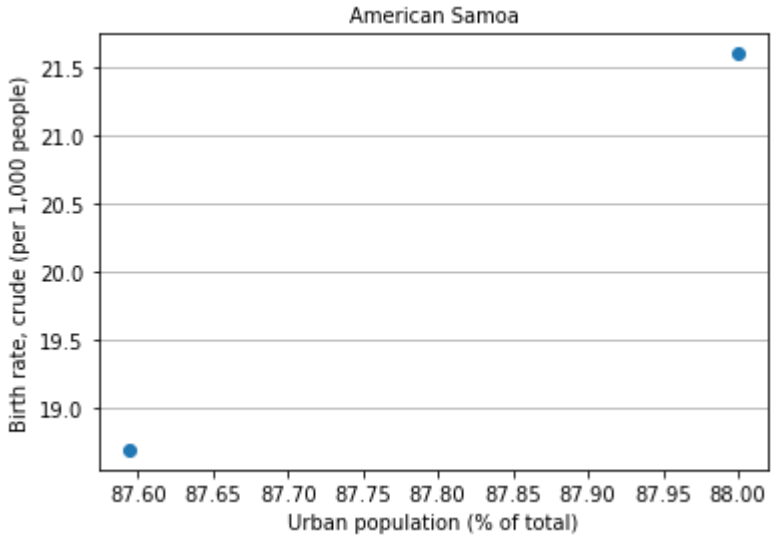

Latin America & Caribbean (developing only)
Birth rate, crude (per 1,000 people)
Urban population (% of total)
45
40
35
30
25
20
45
50
55
60
65
70
75

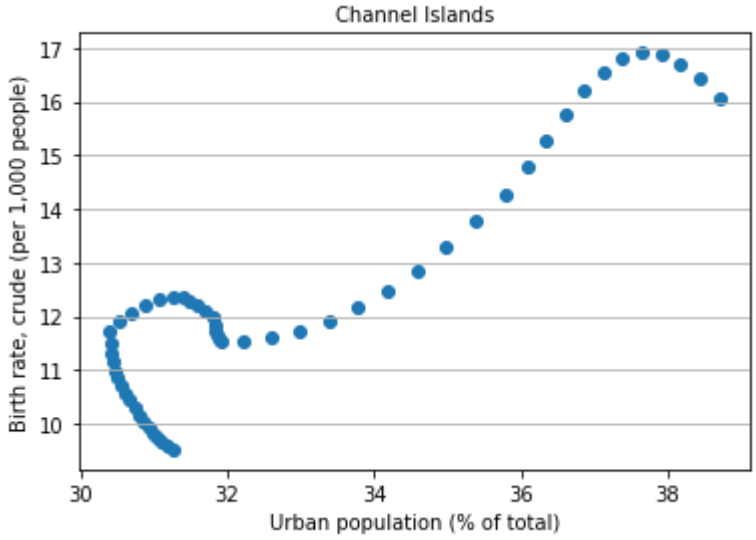

Ecuador
Birth rate, crude (per 1,000 people)
Urban population (% of total)
45
40
35
30
25
20
35
40
45
50
55
60

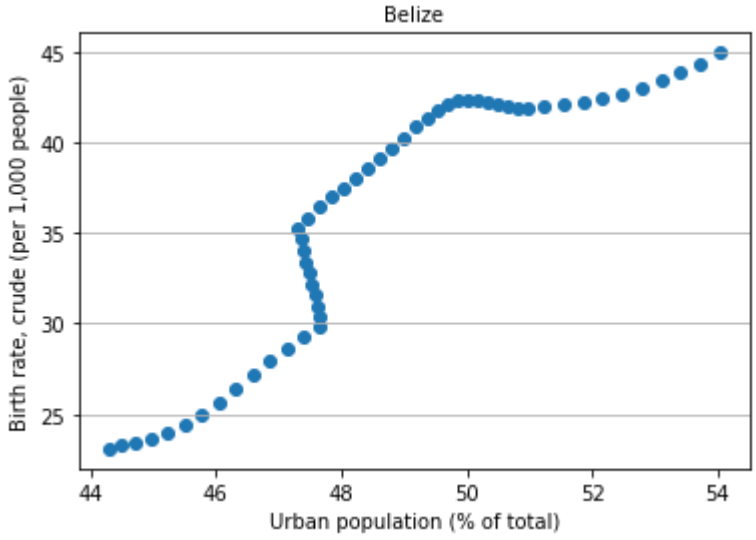

Latin America & Caribbean (all income levels)
Birth rate, crude (per 1,000 people)
Urban population (% of total)
40
35
30
25
20
50
55
60
65
70
75
80

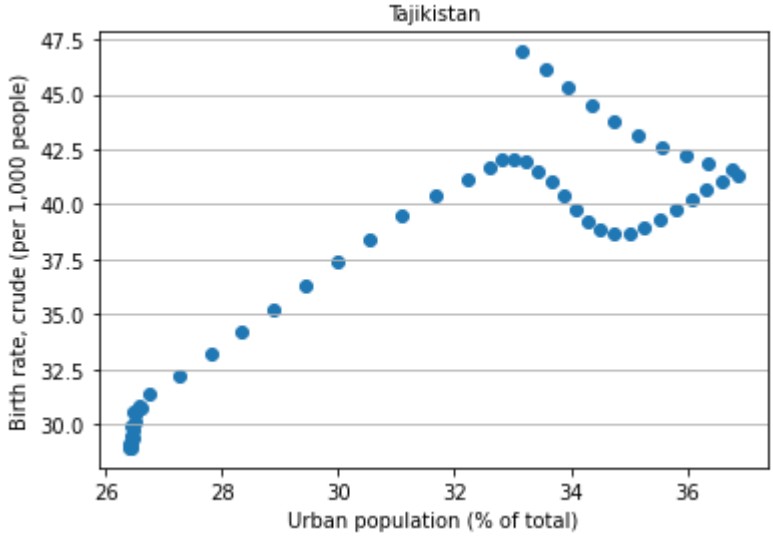
In [97]:

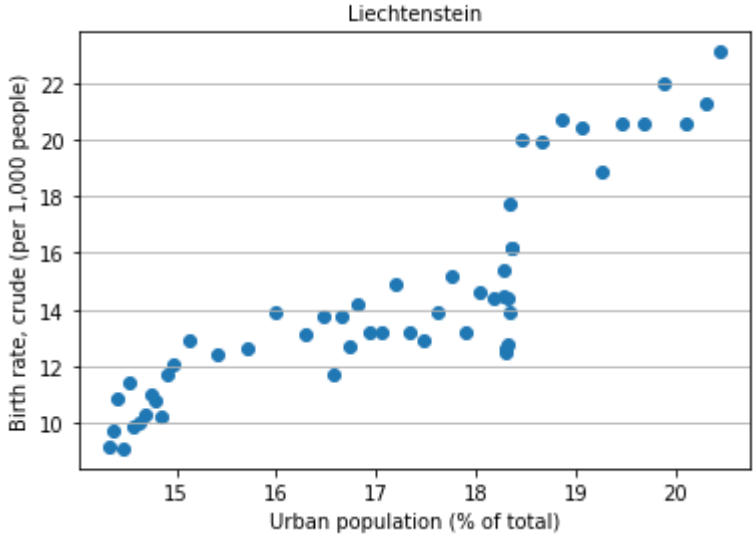
```
for country in lowest_correlation["Country"]:
    country_data = reduced_data[reduced_data["CountryName"] == country]
    country_piv = country_data.pivot(index='Year', columns='IndicatorName', values='Val
ue')
    plot_country(country_piv, country)
```


American Samoa
Birth rate, crude (per 1,000 people)
21.5
21.0
20.5
20.0
19.5
19.0
87.60
87.65
87.70
87.75
87.80
87.85
87.90
87.95
88.00
Urban population (% of total)


Channel Islands
Birth rate, crude (per 1,000 people)
17
16
15
14
13
12
11
10
30
32
34
36
38
Urban population (% of total)


Belize
Birth rate, crude (per 1,000 people)
45
40
35
30
25
44
46
48
50
52
54
Urban population (% of total)


Tajikistan
Birth rate, crude (per 1,000 people)
47.5
45.0
42.5
40.0
37.5
35.0
32.5
30.0
26
28
30
32
34
36
Urban population (% of total)


Liechtenstein
Birth rate, crude (per 1,000 people)
22
20
18
16
14
12
10
15
16
17
18
19
20
Urban population (% of total)