World development indicators

Which country will develop more

Group 12:
Stefano Moawad
Leonardo Comandini
Diana Isaeva
Andrea Schiavon
Viktor Snesarevskii

April-May 2017

Our Goals

Abstract: “. . . individuate the best countries where to invest, w.r.t. different economical and social fields"

- Best = ?

Our Goals

Abstract: "... individuate the best countries where to invest, w.r.t. different economical and social fields"

- Best = upcoming largest development

Our Goals

Abstract: “. . . individuate the best countries where to invest, w.r.t. different economical and social fields"

- Best = upcoming largest development
- Development = ?

Our Goals

Abstract: "... individuate the best countries where to invest, w.r.t. different economical and social fields"

- Best = upcoming largest development
- Development = ?
- Invest = ?

Raw dataset

Indicators (5 656 458 x 6)

Country name | Country code | Indicator name | Indicator code | Year | Value

Country (247 x 31)

Country code | Short name | Table name | Long name | Alpha 2 code

Currency unit | Special notes | Region | Indice group | **etc...**

Country notes (4 857 x 3)

Country code | Series code | Description

Series (1 345 x 20)

Series code | Topic | Indicator name | Short definition

Long definition | Unit of measure | Periodicity | **etc...**

Series notes (369 x 3)

Series code | Year | Description

Raw dataset

Indicators (5 656 458 x 6)

Country name | Country code | Indicator name | Indicator code | Year | Value

Country (247 x 31)

Country code | Short name | Table name | Long name | Alpha 2 code

Currency unit | Special notes | Region | Indice group | **etc...**

Country notes (4 857 x 3)

Country code | Series code | Description

Series (1 345 x 20)

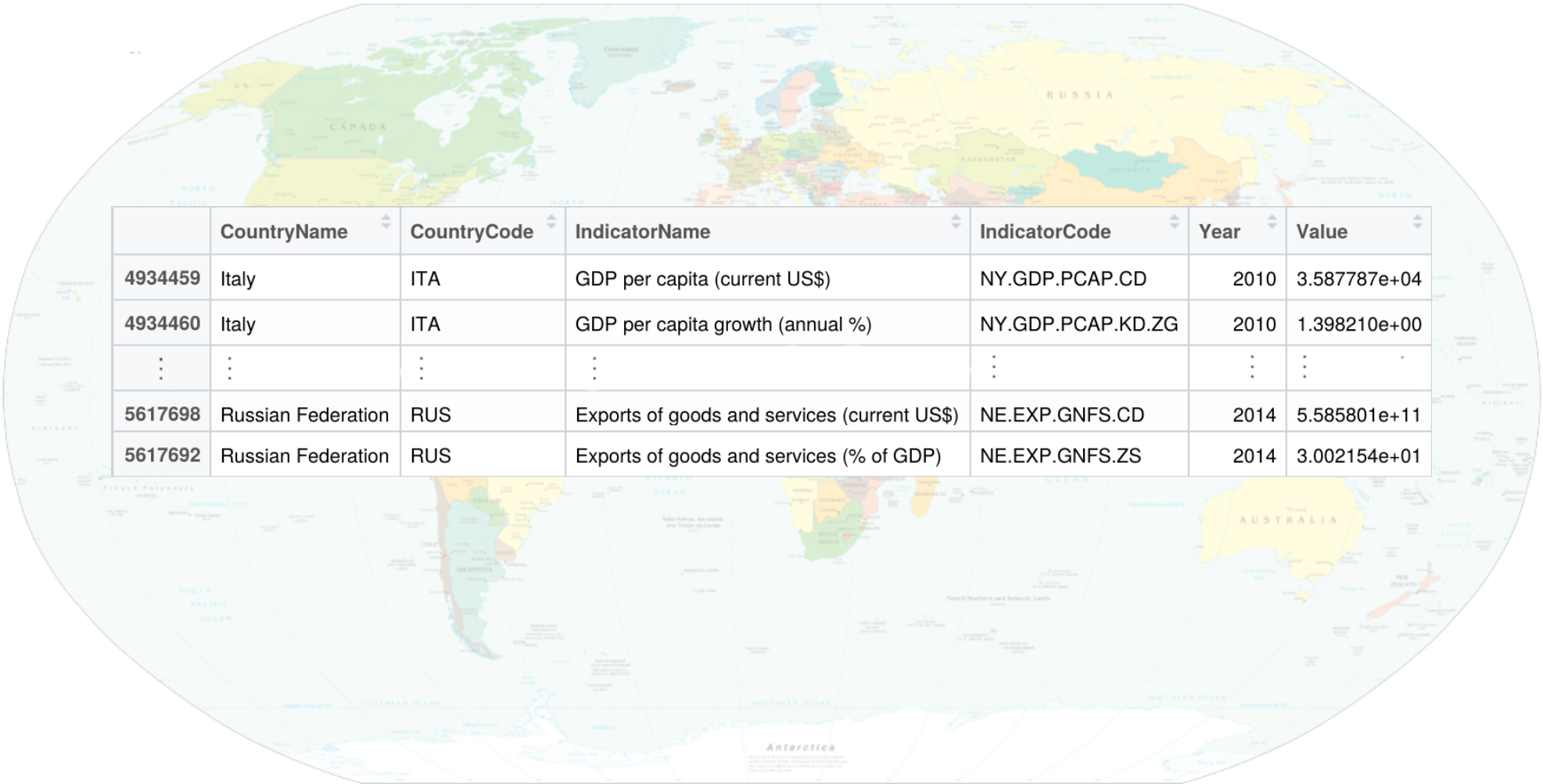
Series code | Topic | Indicator name | Short definition

Long definition | Unit of measure | Periodicity | **etc...**

Series notes (369 x 3)

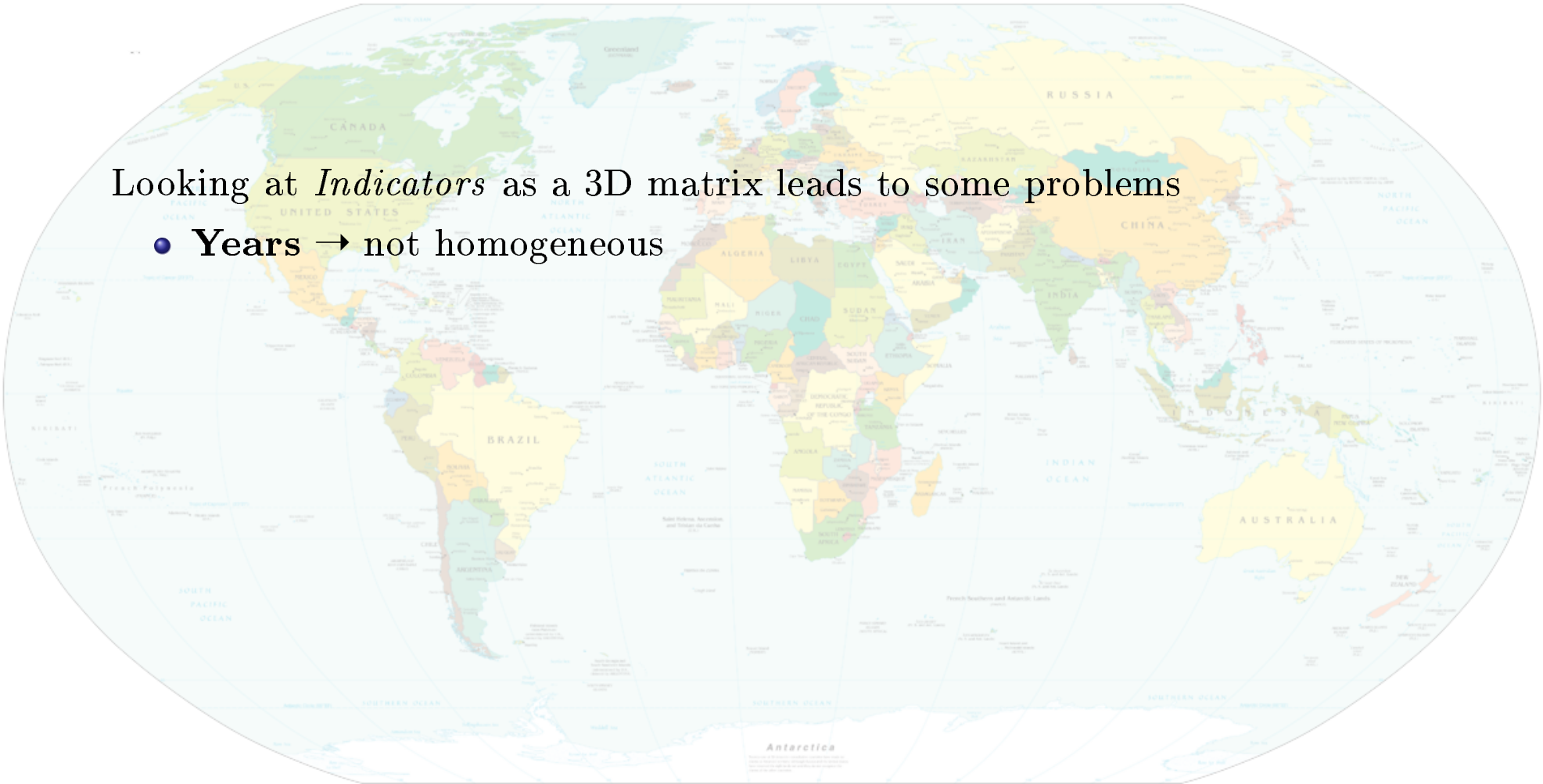
Series code | Year | Description

Examples of indicators



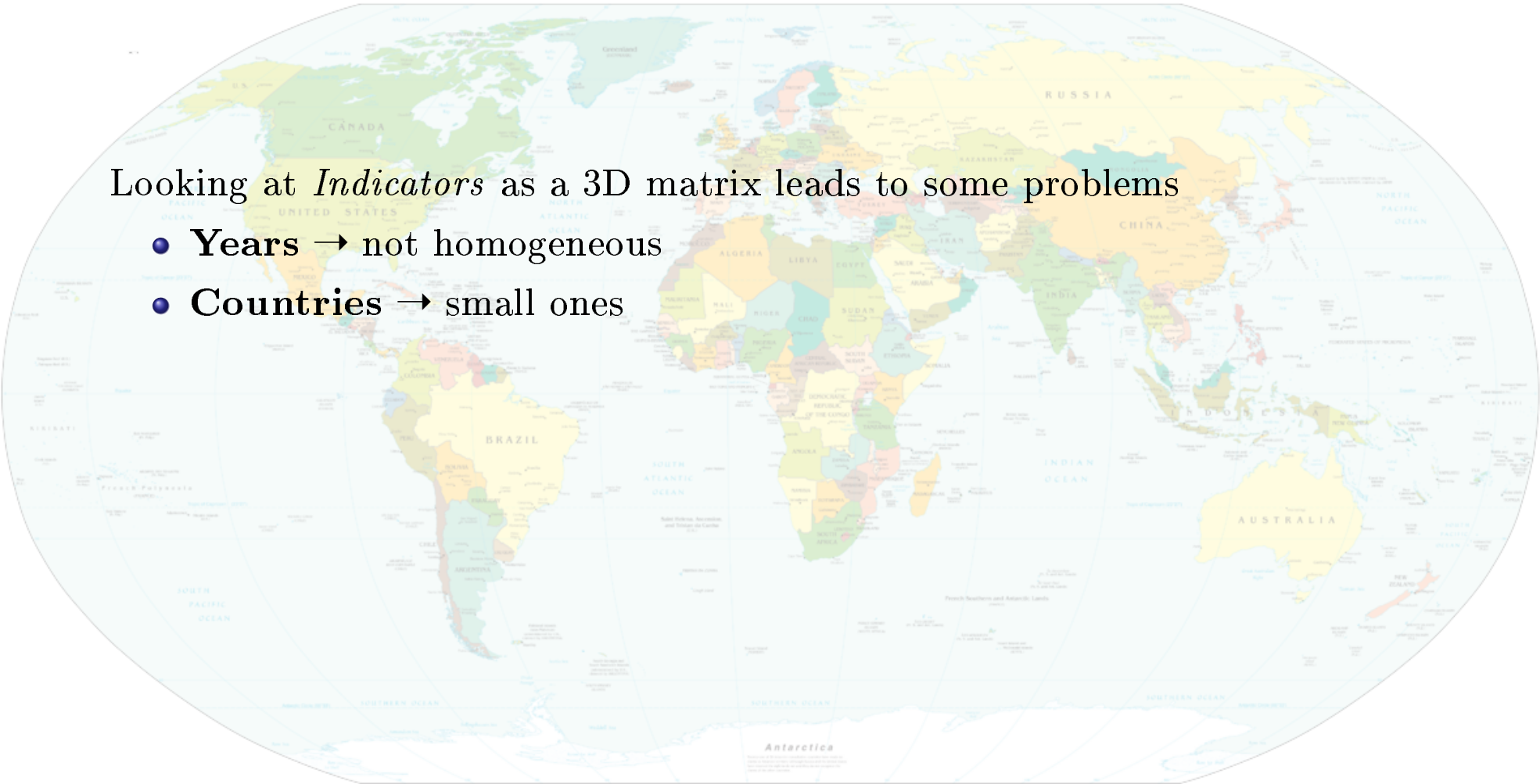
	CountryName	CountryCode	IndicatorName	IndicatorCode	Year	Value
4934459	Italy	ITA	GDP per capita (current US$)	NY.GDP.PCAP.CD	2010	3.587787e+04
4934460	Italy	ITA	GDP per capita growth (annual %)	NY.GDP.PCAP.KD.ZG	2010	1.398210e+00
⋮	⋮	⋮	⋮	⋮	⋮	⋮
5617698	Russian Federation	RUS	Exports of goods and services (current US$)	NE.EXP.GNFS.CD	2014	5.585801e+11
5617692	Russian Federation	RUS	Exports of goods and services (% of GDP)	NE.EXP.GNFS.ZS	2014	3.002154e+01

	CountryName	Access to electricity (% of population)	Agricultural land (sq. km)	Net income from abroad (current US$)	Urban population	. . . .
1	Canada	100.00000	677680	-21448405896	21282904	
2	Cuba	92.86102	67410	-610497598	7763439	
3	Italy	100.00000	168400	-15262429218	37846480	
4	Japan	100.00000	56930	20113271442	95542280	
5	Mongolia	79.81566	1256560	-43600000	1245683	
⋮						



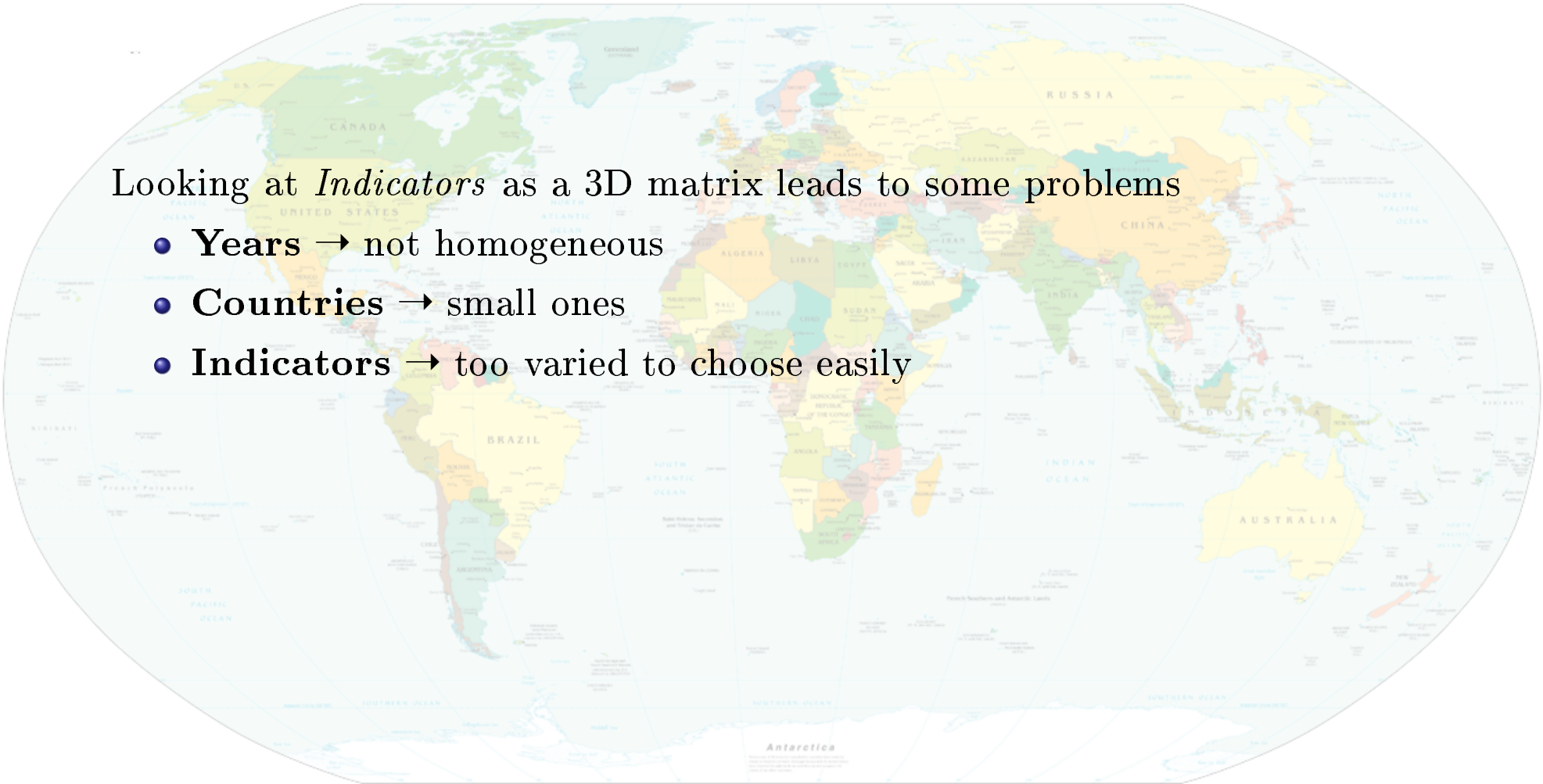
Looking at *Indicators* as a 3D matrix leads to some problems

- **Years** → not homogeneous



Looking at *Indicators* as a 3D matrix leads to some problems

- **Years** → not homogeneous
- **Countries** → small ones

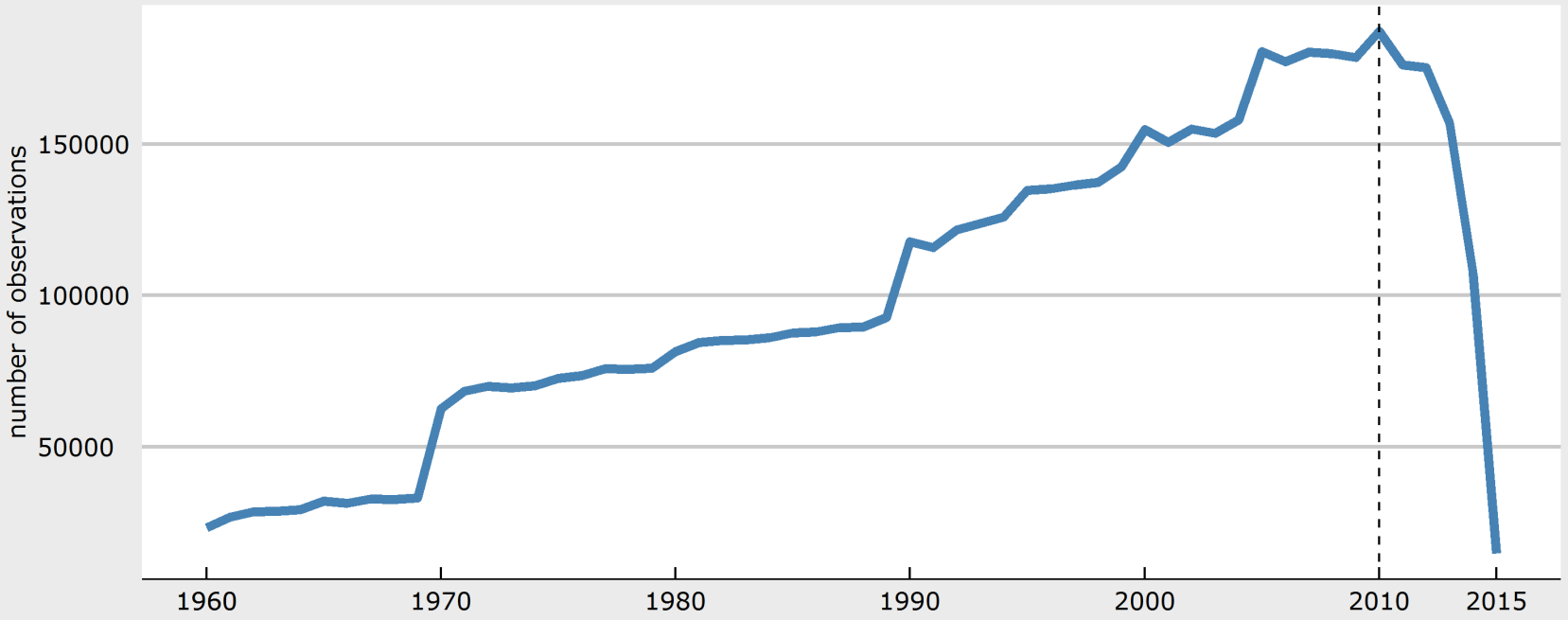


Looking at *Indicators* as a 3D matrix leads to some problems

- **Years** → not homogeneous
- **Countries** → small ones
- **Indicators** → too varied to choose easily

Number of observations is not homogeneous

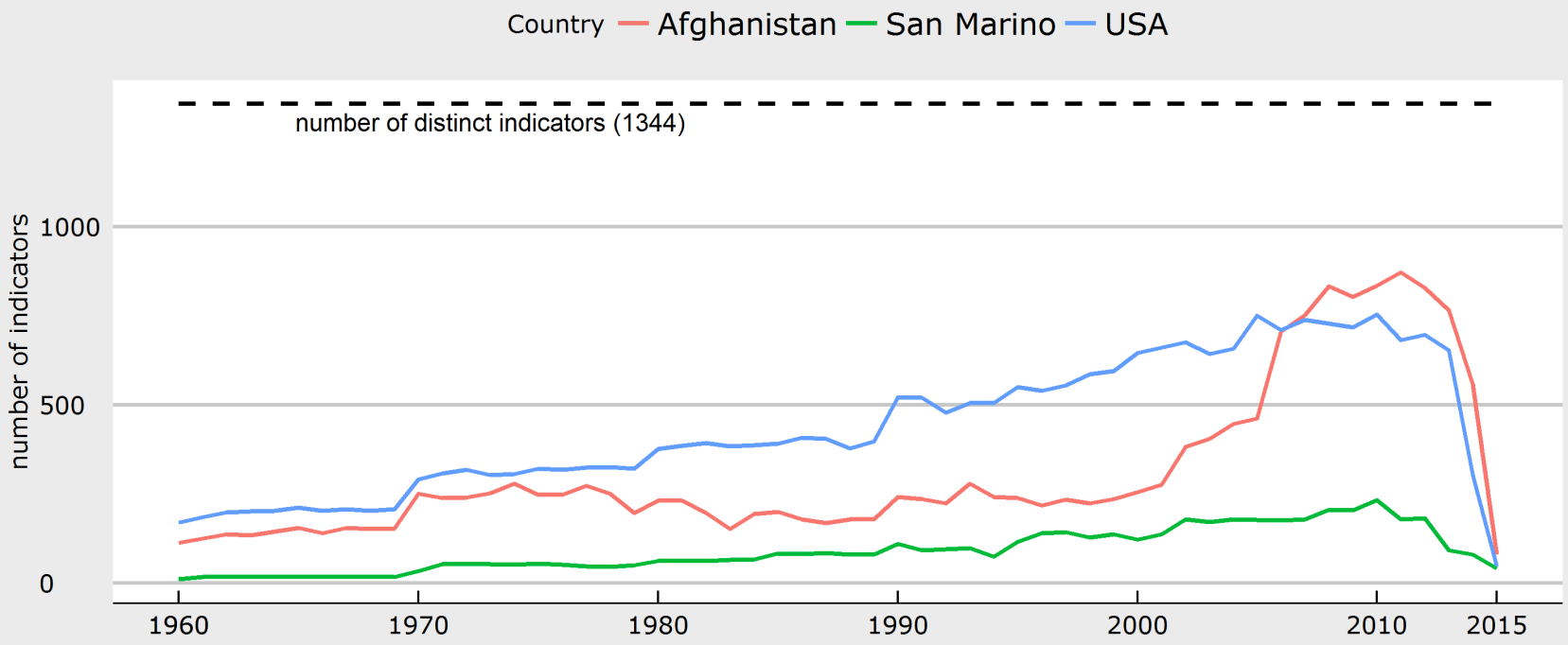
Increasing in the years, dropping in the very last due to the difficulty to find recent data



For each year, an indicator is counted multiple times if present for multiple countries

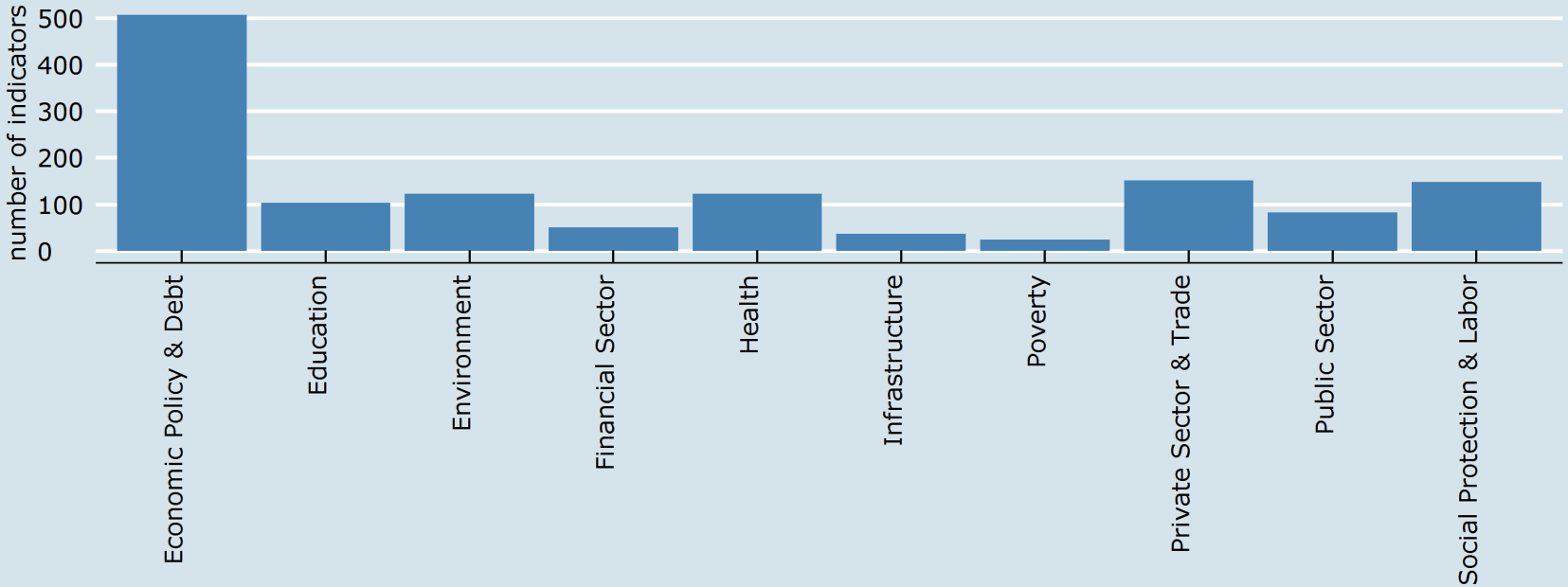
Three different countries behaviour

USA are normal, San Marino is small, Afghanistan is problematic



Number of indicators for each supertopic

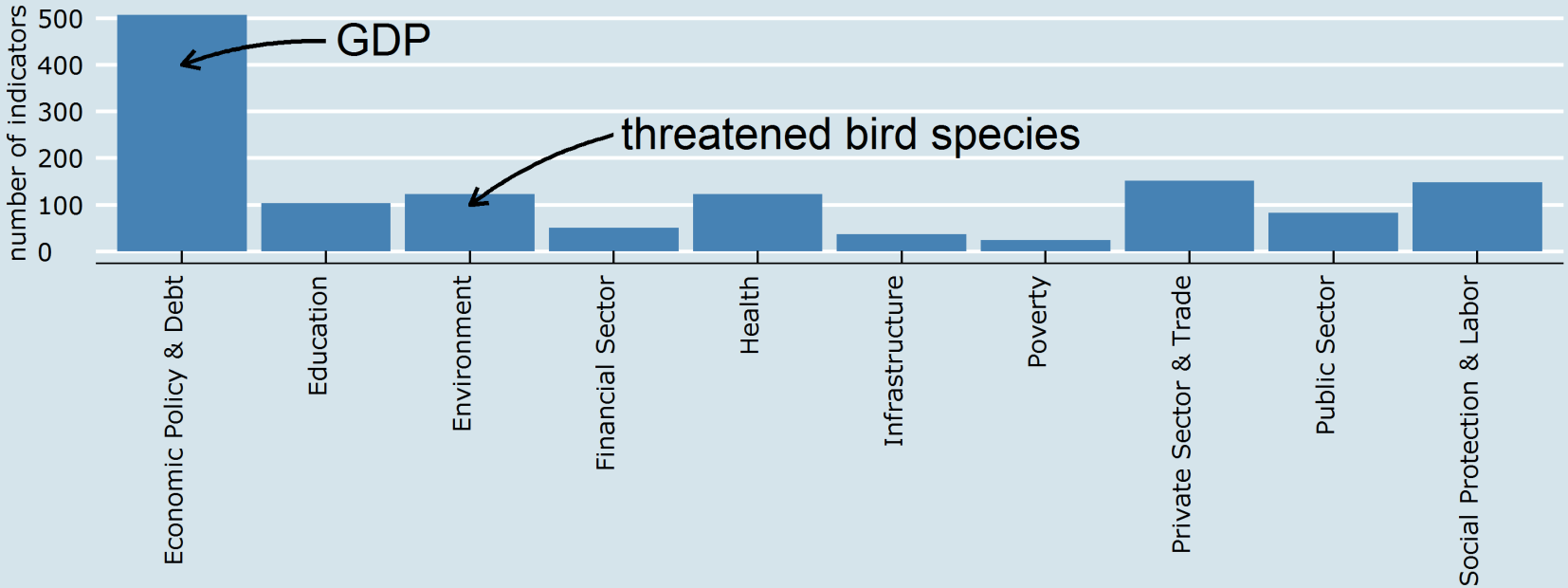
Prevalence of economic indicators: great for our analysis



The supertopic is not present in the raw dataset but comes naturally from the topic.
(e.g. topic = Infrastructure: Transportation, supertopic = Infrastructure)

Number of indicators for each supertopic

Prevalence of economic indicators: great for our analysis



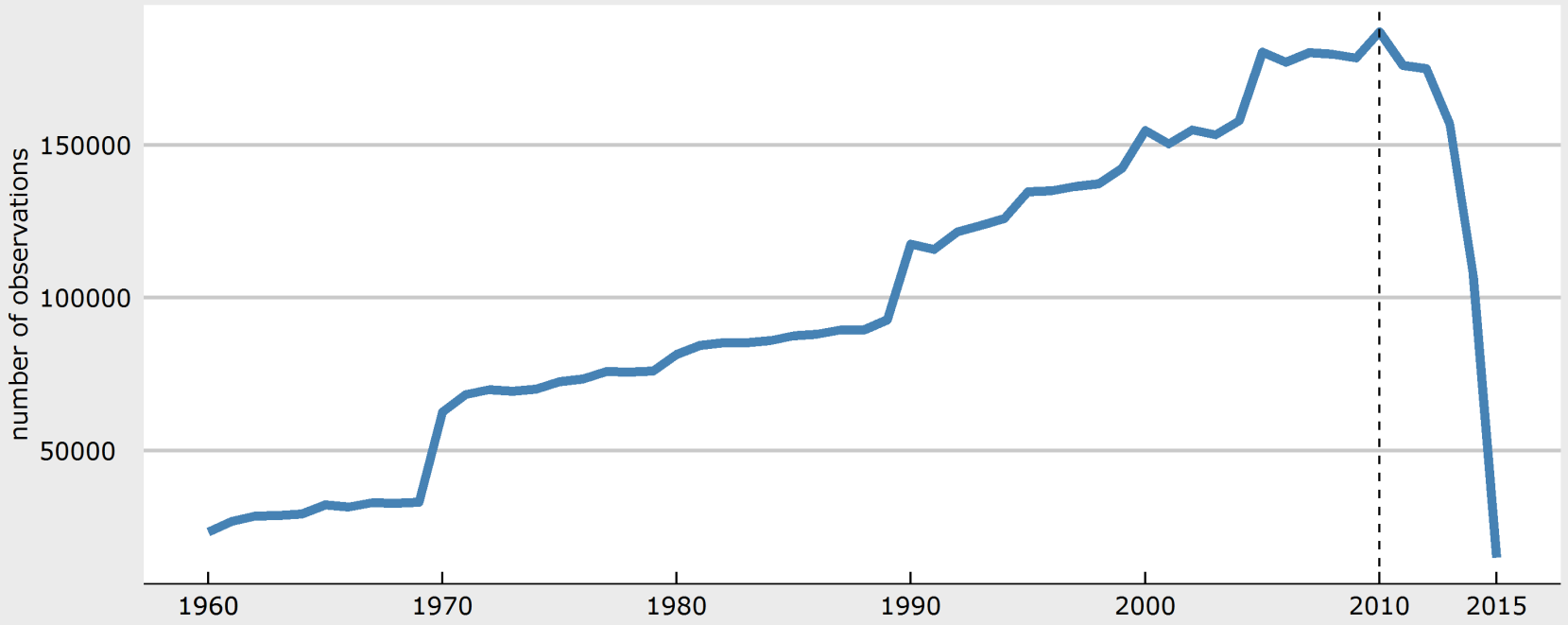
The supertopic is not present in the raw dataset but comes naturally from the topic.
(e.g. topic = Infrastructure: Transportation, supertopic = Infrastructure)

An intermediate goal:

Extract a full matrix with meaningful indicators to perform PCA

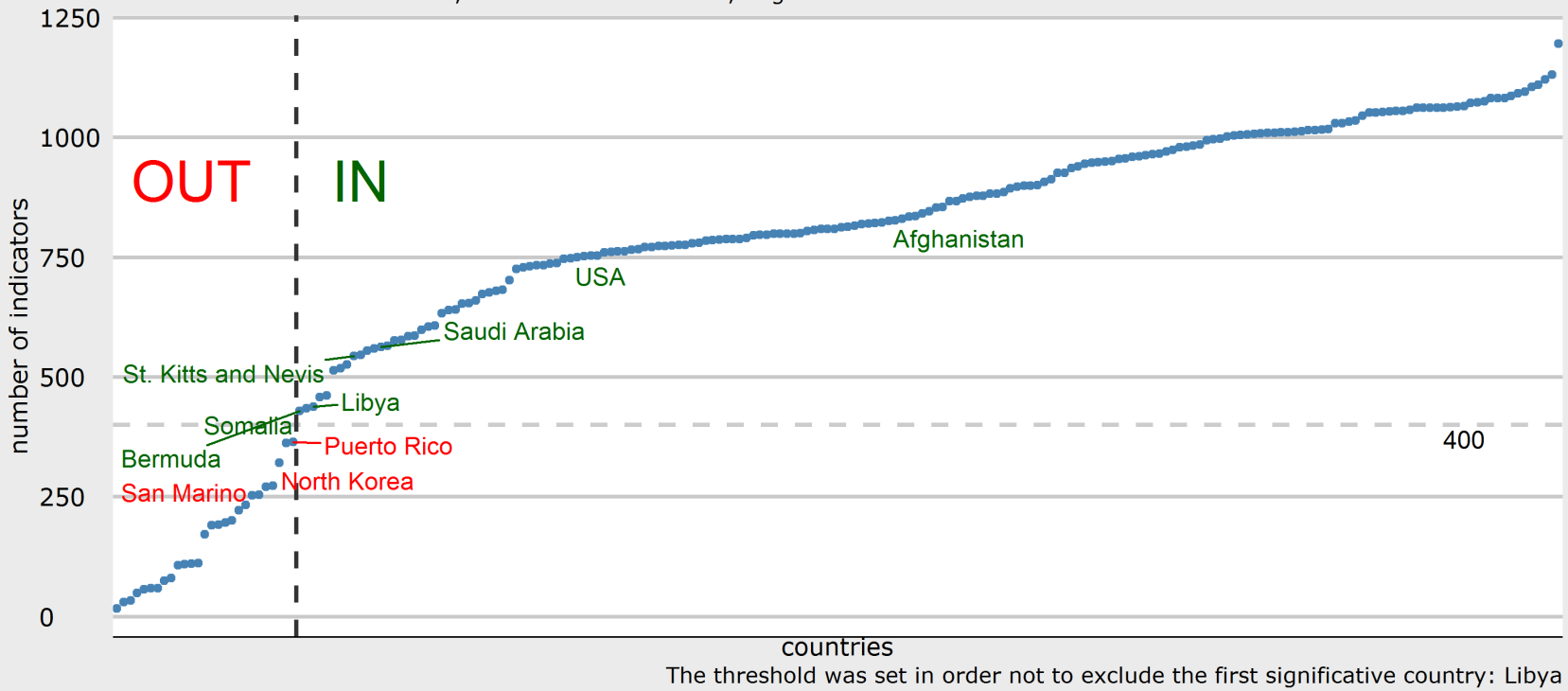
1. Fix the best year: 2010

Put aside the year dipendency for simplicity



2. Shrink the countries with too few indicators

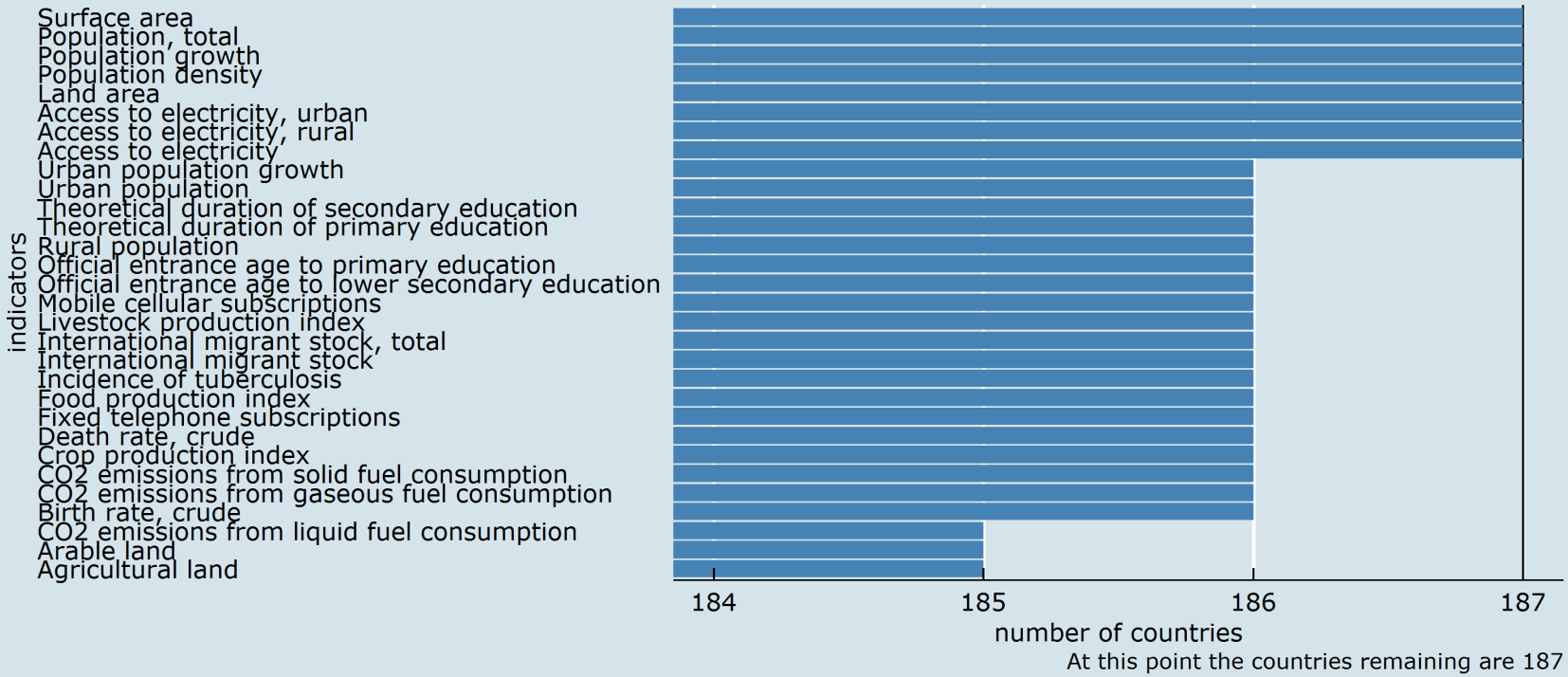
Threshold set at 400 indicators, in red the countries out, in green the countries in



The threshold was set in order not to exclude the first significative country: Libya

3. Select the indicators among the fullest

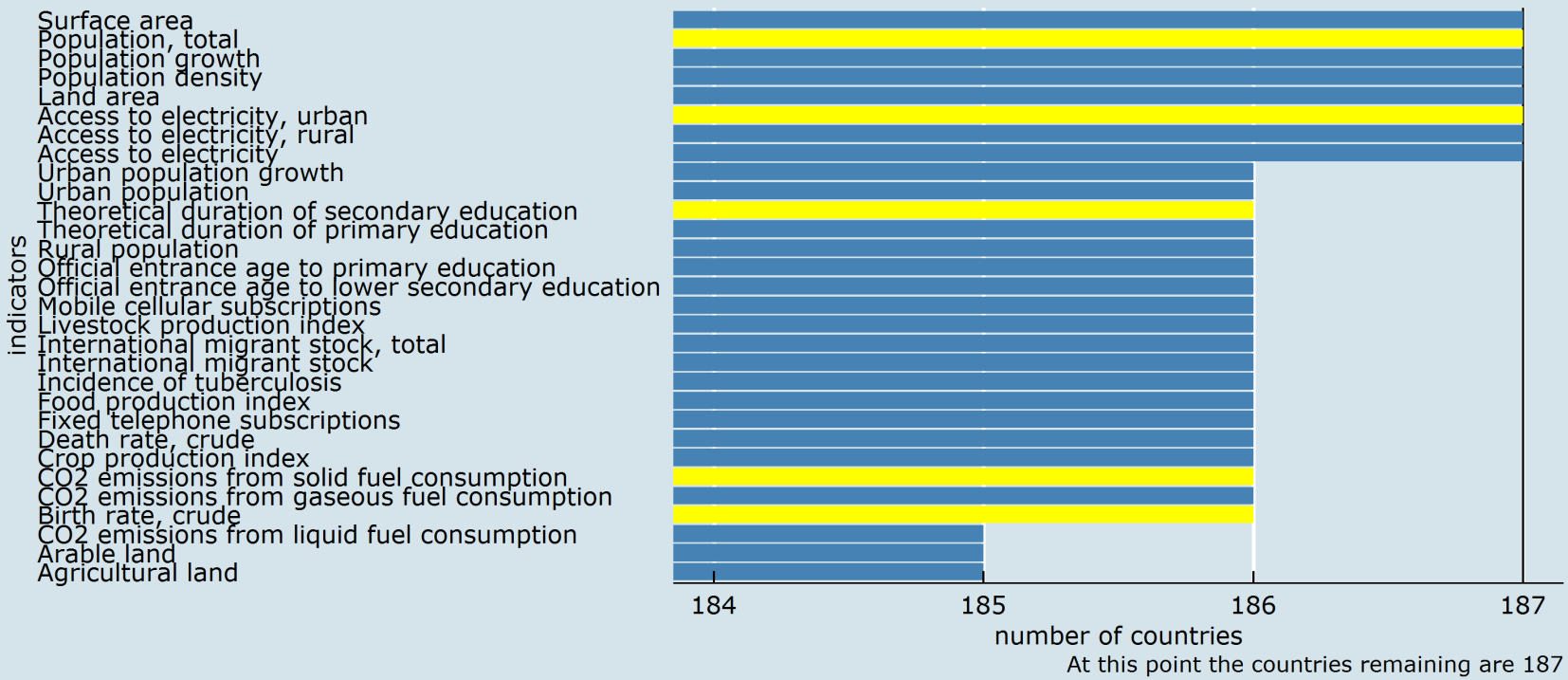
Look at the indicators manually and select the most significative



At this point the countries remaining are 187

3. Select the indicators among the fullest

Look at the indicators manually and select the most significative



At this point the countries remaining are 187

Next steps:

4. Evaluate the fullness of the matrix

Next steps:

4. Evaluate the fullness of the matrix
5. If problematic countries or indicators are still present, consider shrinking them

Next steps:

4. Evaluate the fullness of the matrix
5. If problematic countries or indicators are still present, consider shrinking them
6. Fill the *real* missing data (via interpolation, value at the previous year or manually from the source)

Next steps:

4. Evaluate the fullness of the matrix
5. If problematic countries or indicators are still present, consider shrinking them
6. Fill the *real* missing data (via interpolation, value at the previous year or manually from the source)
7. Expand the shrunk matrix over years

Example 1

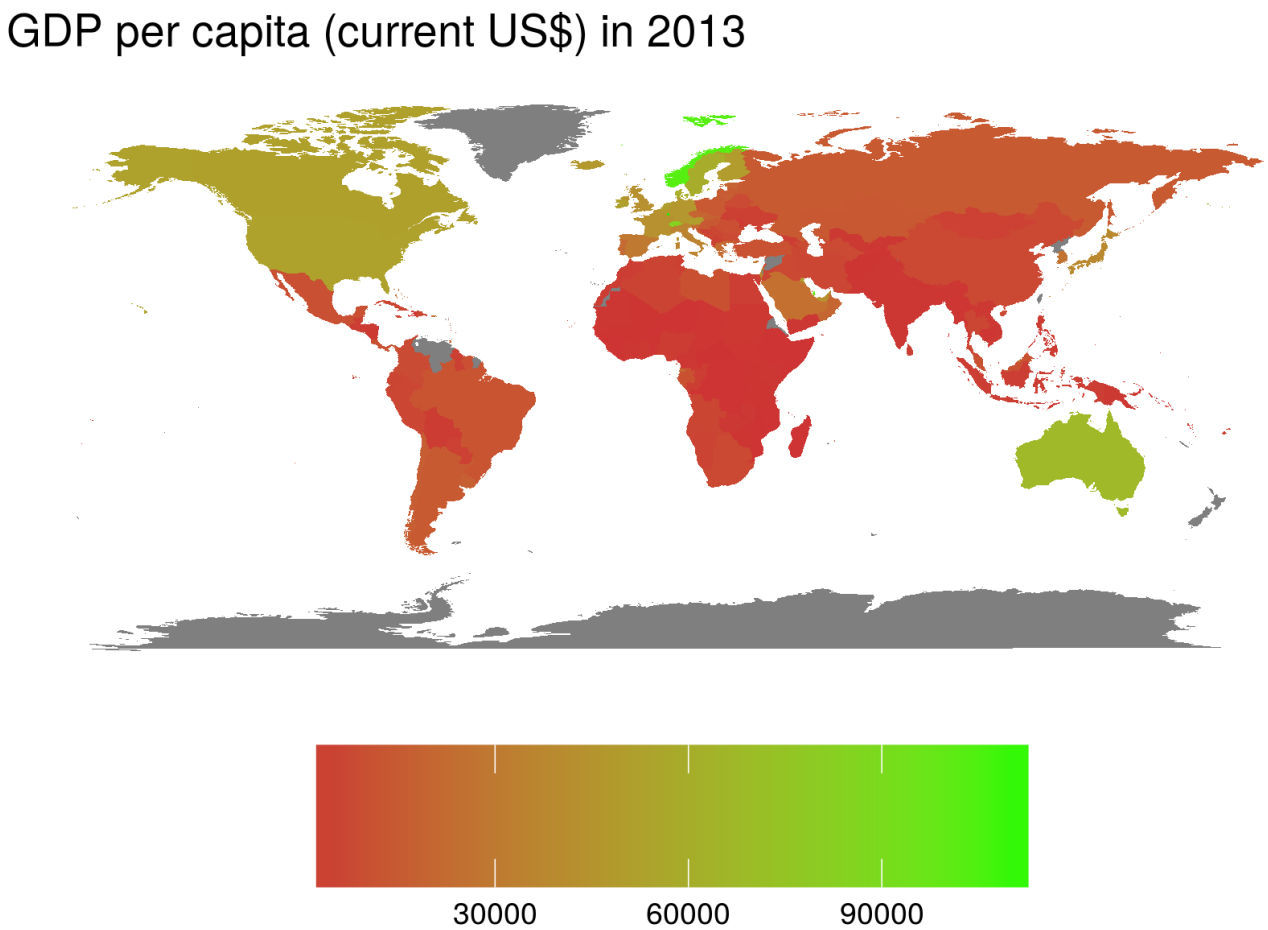


Рис.: GDP per capita, 2013

Example 2

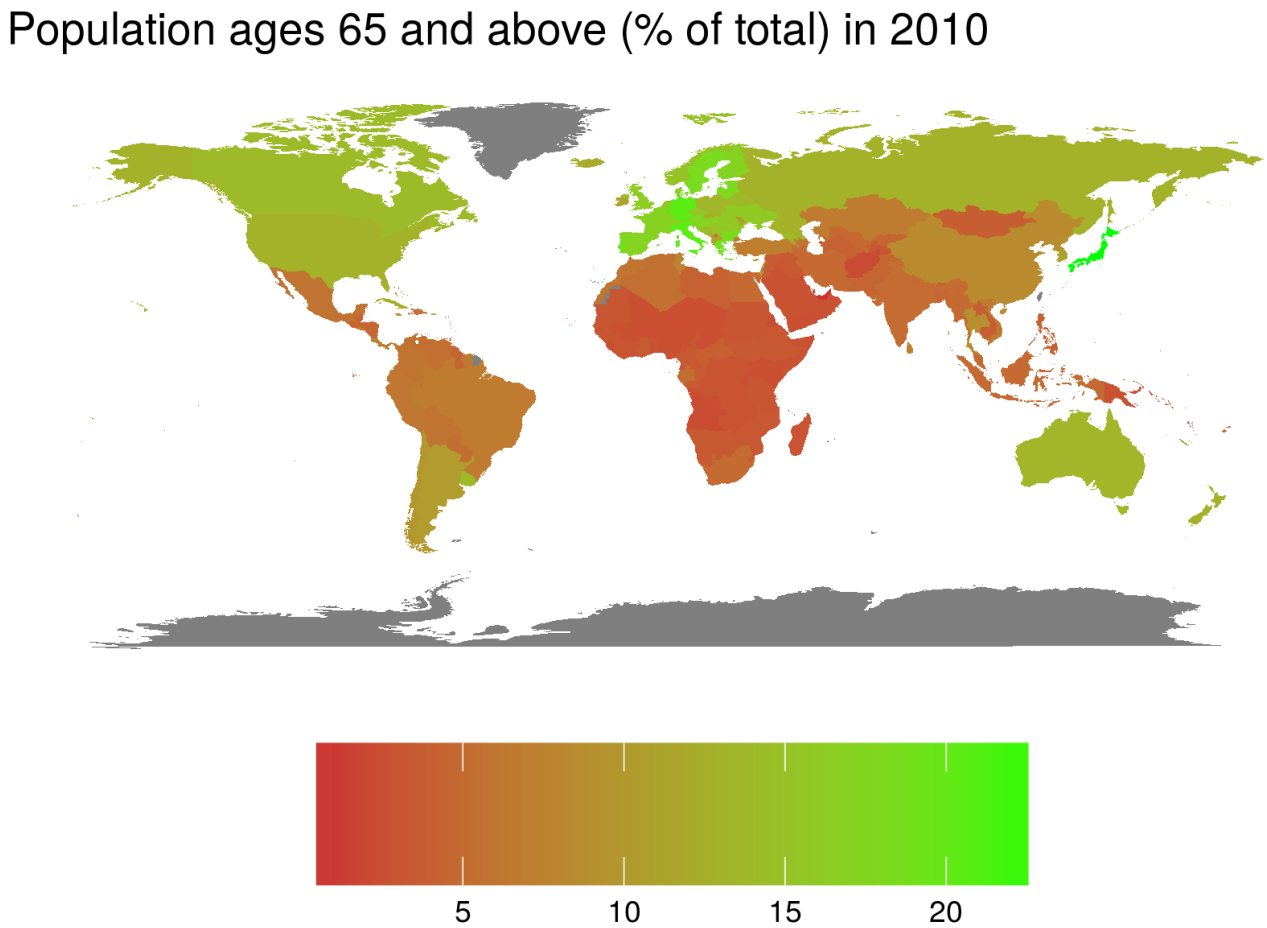


Рис.: Population aged over 65

Example 3

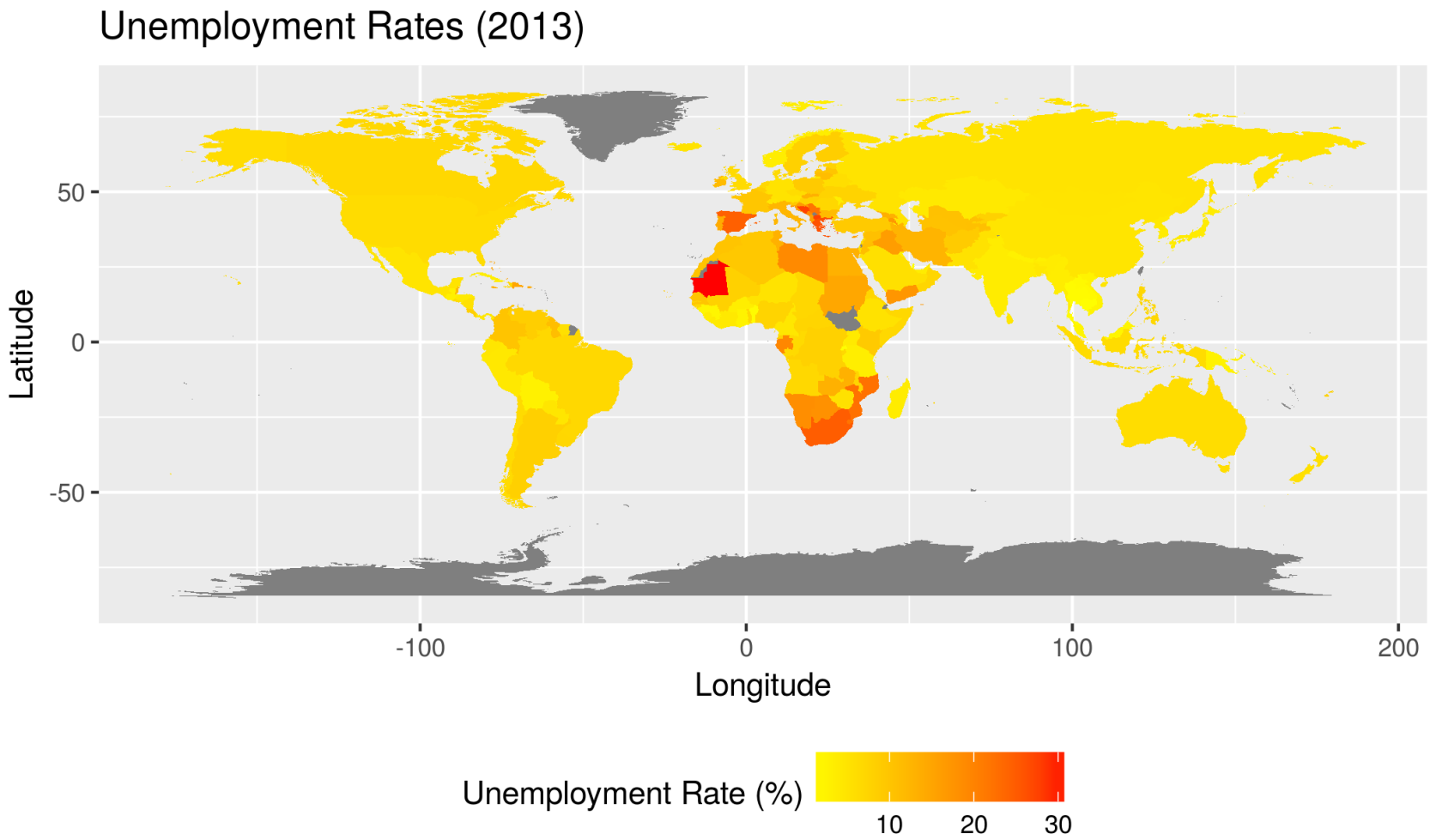


Рис.: World unemployment