

# World development indicators

Which country will develop more

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

April-May 2017

kaggle



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

- best = ?



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

- best = ?
- best = upcoming largest development

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

- best = ?
- best = upcoming largest development
- development = ?

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

- best = ?
- best = upcoming largest development
- development = ?
- Invest = ?

# Raw dataset

## Indicators (6 columns)

Country name	Country code	Indicator name	Indicator code	Year	Value
--------------	--------------	----------------	----------------	------	-------

## Country (31 columns)

Country code	Short name	Table name	Long name	Alpha 2 code
Currency unit	Special notes	Region	Indice group	etc...

## Country notes (3 columns)

Country code	Series code	Description
--------------	-------------	-------------

## Series (20 columns)

Series code	Topic	Indicator name	Short definition
Long definition	Unit of measure	Periodicity	etc...

## Series notes (3 columns)

Series code	Year	Description
-------------	------	-------------

# Raw dataset

## Indicators (6 columns)

Country name	Country code	Indicator name	Indicator code	Year	Value
--------------	--------------	----------------	----------------	------	-------

## Country (31 columns)

Country code	Short name	Table name	Long name	Alpha 2 code
Currency unit	Special notes	Region	Indice group	etc...

## Country notes (3 columns)

Country code	Series code	Description
--------------	-------------	-------------

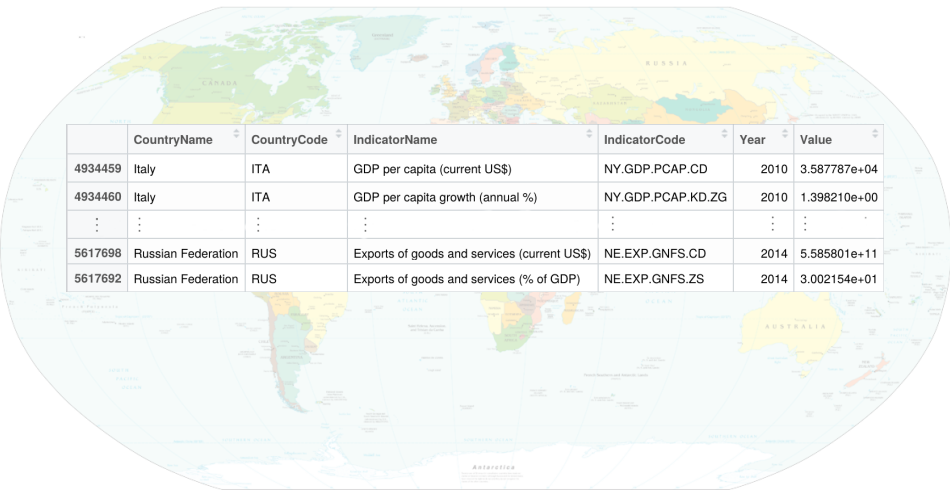
## Series (20 columns)

Series code	Topic	Indicator name	Short definition
Long definition	Unit of measure	Periodicity	etc...

## Series notes (3 columns)

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

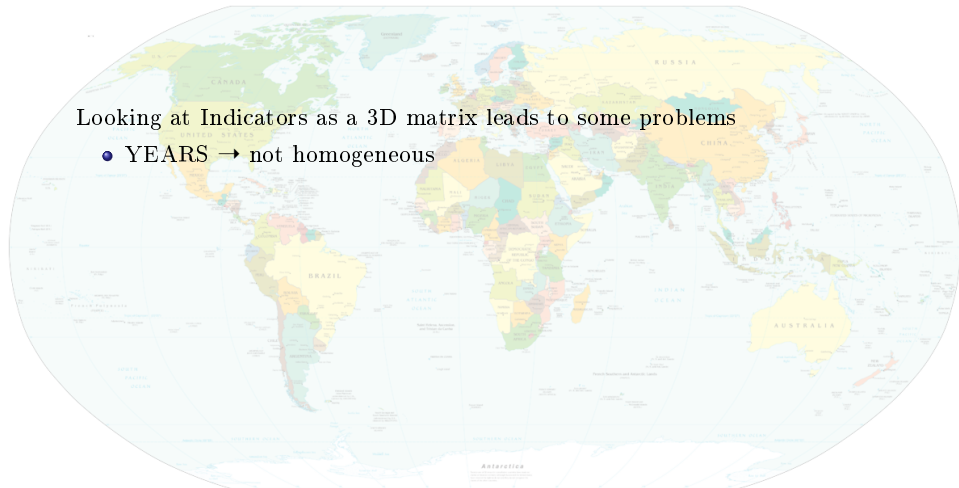


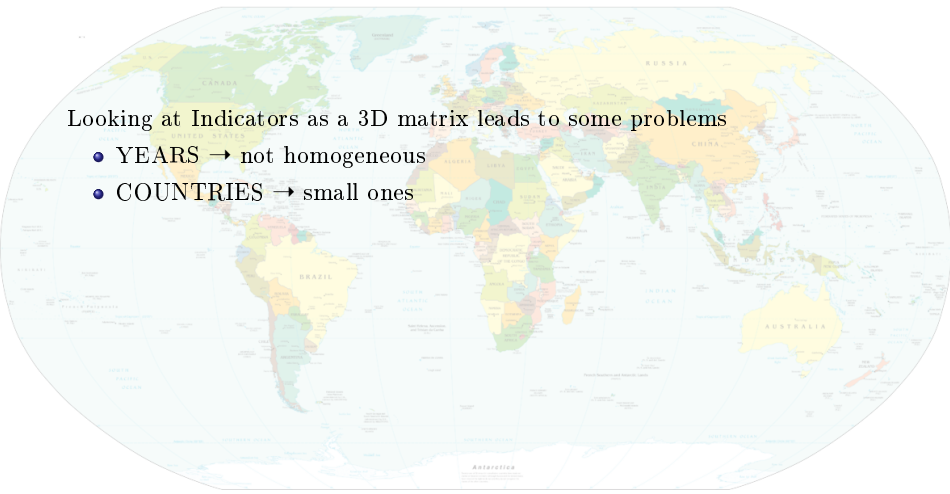
	Countrycode	Seriescode	Description
1	ABW	EG.EGY.PRIM.PP.KD	Sources: Estimated based on UN Energy Statistics (2014...
2	ABW	EG.ELC.RNEW.ZS	Sources: UN Energy Statistics (2014)
3	ABW	EG.FEC.RNEW.ZS	Sources: UN Energy Statistics (2014)

			CountryName	CountryCode	IndicatorName	IndicatorCode	Year	Value		
4	ABW	SM.POP	5415624	Bulgaria	BGR	Portfolio equity, net inflows (BoP, current US\$)	BX.PEF.TOTL.CD.WD	2013	-1.910000e+07	
5	ABW	SP.DYN	5415625	Bulgaria	BGR	Portfolio investment, bonds (PPG + PNG) (NFL, current ...	DT.NFL.BOND.CD	2013	6.344960e+08	
6	ABW	SP.DYN	5415626	Bulgaria	BGR	Portfolio investment, net (BoP, current US\$)	BN.KLT.PTXL.CD	2013	1.543552e+08	
7	ABW	SP.DYN	5415627	Bulgaria	BGR	Poverty headcount ratio at national poverty lines (% of po...	SI.POV.NAHC	2013	2.180000e+01	
8	ABW	SP.DYN	5415628	Bulgaria	BGR	Power outages in firms in a typical month (number)	IC.ELC.OUTG	2013	1.200000e+00	
9	ABW	SP.DYN	5415629	Bulgaria	BGR	PPG, bilateral (AMT, current US\$)	DT.AMT.BLAT.CD	2013	6.899600e+07	
10	ABW	SP.DYN	5415630	Bulgaria	BGR	PPG, bilateral (DIS, current US\$)	DT.DIS.BLAT.CD	2013	9.600000e+04	
11	ABW	SP.DYN	5415631	Bulgaria	BGR	PPG, bilateral (DOD, current US\$)	DT.DOD.BLAT.CD	2013	4.888450e+08	
12	ABW	SP.DYN	5415632	Bulgaria	BGR	PPG, bilateral (INT, current US\$)	DT.INT.BLAT.CD	2013	1.514000e+07	
13	ABW	SP.DYN	5415633	Bulgaria	BGR	PPG, bilateral (NTR, current US\$)	DT.NTR.BLAT.CD	2013	-8.404000e+07	Nations World Population Prospects
			5415634	Bulgaria	BGR	PPG, bilateral (TDS, current US\$)	DT.TDS.BLAT.CD	2013	8.413600e+07	Nations World Population Prospects
			5415635	Bulgaria	BGR	PPG, bilateral concessional (AMT, current US\$)	DT.AMT.BLTC.CD	2013	6.631000e+07	Nations World Population Prospects
			5415636	Bulgaria	BGR	PPG, bilateral concessional (DIS, current US\$)	DT.DIS.BLTC.CD	2013	9.600000e+04	Nations World Population Prospects
			5415637	Bulgaria	BGR	PPG, bilateral concessional (DOD, current US\$)	DT.DOD.BLTC.CD	2013	4.736750e+08	Nations World Population Prospects
			5415638	Bulgaria	BGR	PPG, bilateral concessional (INT, current US\$)	DT.INT.BLTC.CD	2013	1.410900e+07	Nations World Population Prospects
			5415639	Bulgaria	BGR	PPG, bilateral concessional (NFL, current US\$)	DT.NFL.BLTC.CD	2013	-6.621400e+07	and female life expectancy
			5415640	Bulgaria	BGR	PPG, bilateral concessional (NTR, current US\$)	DT.NTR.BLTC.CD	2013	-8.032300e+07	Nations World Population Prospects
			5415641	Bulgaria	BGR	PPG, bilateral concessional (TDS, current US\$)	DT.TDS.BLTC.CD	2013	8.041900e+07	Nations World Population Prospects
				</						

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

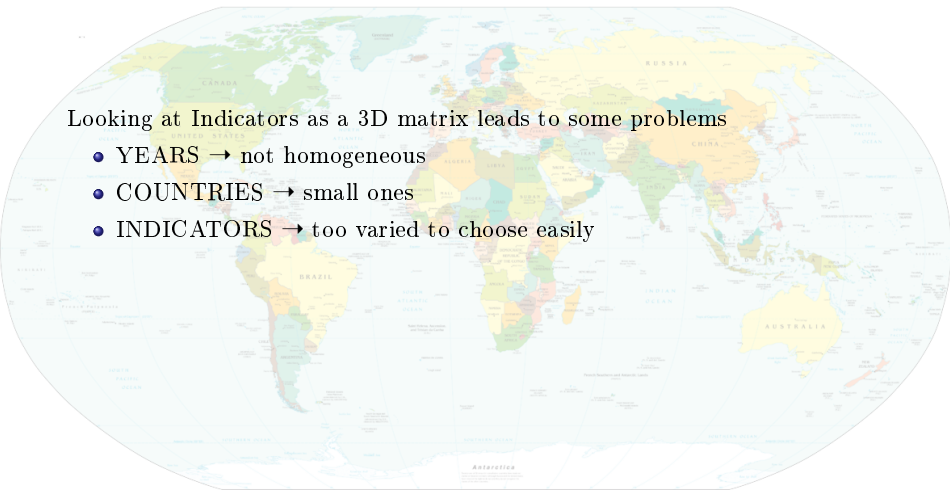
- YEARS → not homogeneous





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

- YEARS → not homogeneous
- COUNTRIES → small ones

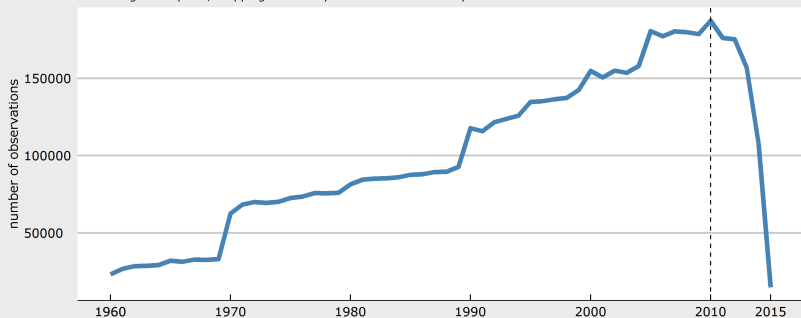


Looking at Indicators as a 3D matrix leads to some problems

- 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

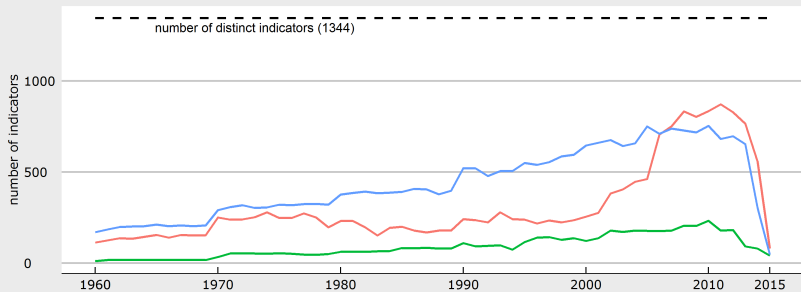


For each year, an indicator is counted multiple times if present for multiple countries

### Three different countries behaviour

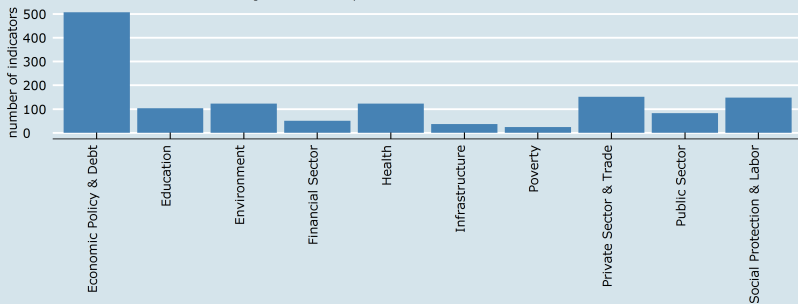
USA are normal, San Marino is small, Afghanistan is problematic

Country — Afghanistan — San Marino — USA



## Number of indicators for each supertopic

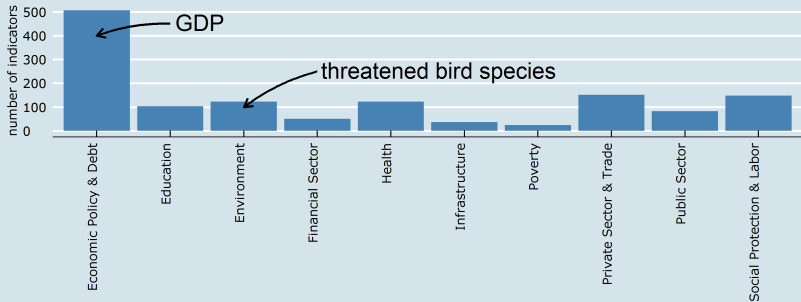
Prevalence of economic indicators: great for our analysis



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)

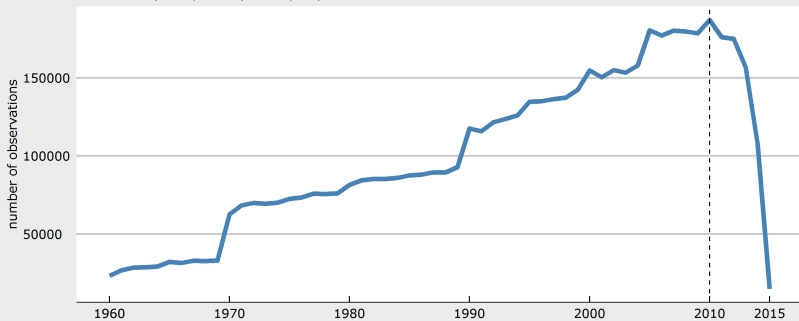




An intermediate goal:  
Extract a full matrix with meaningful indicators to perform PCA

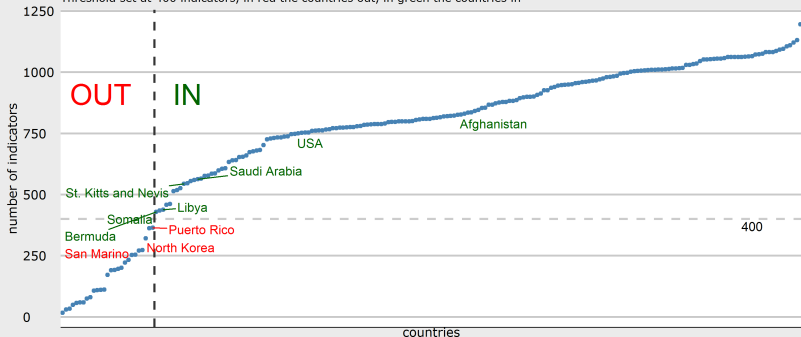
## 1. Fix the best year: 2010

Put aside the year dependency for simplicity



## 2. Shrink the countries with too few indicators

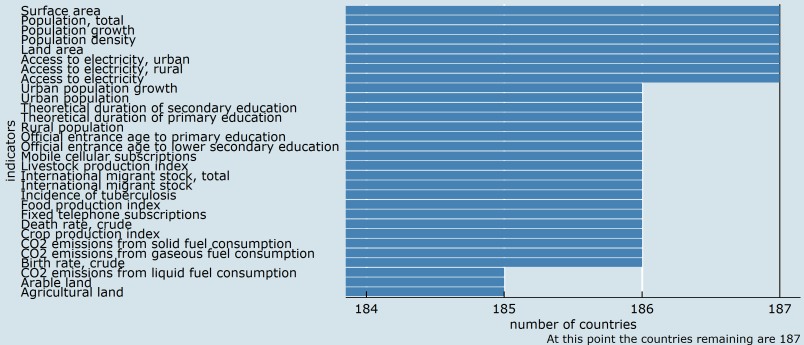
Threshold set at 400 indicators, in red the countries out, in green the countries in



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

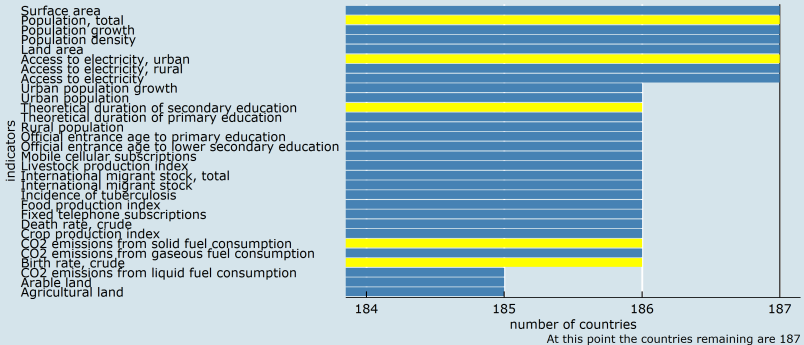
### 3. Select the indicators among the fullest

Look at the indicators manually and select the most significant



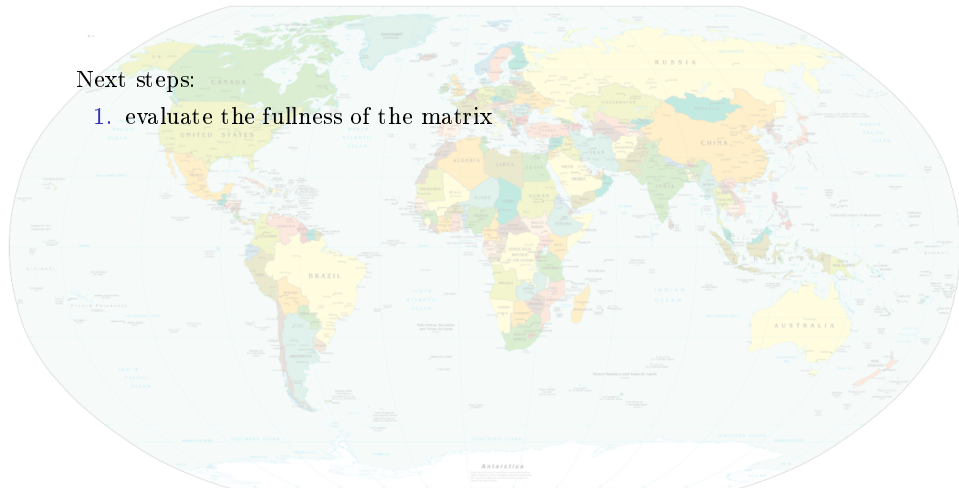
### 3. Select the indicators among the fullest

Look at the indicators manually and select the most significant



Next steps:

1. evaluate the fullness of the matrix



Next steps:

1. evaluate the fullness of the matrix
2. if problematic countries or indicators are still present, consider shrinking them

Next steps:

1. evaluate the fullness of the matrix
2. if problematic countries or indicators are still present, consider shrinking them
3. fill the real missing data (via interpolation, value at the previous year or manually from the source)



Next steps:

1. evaluate the fullness of the matrix
2. if problematic countries or indicators are still present, consider shrinking them
3. fill the real missing data (via interpolation, value at the previous year or manually from the source)
4. expand the shrunk matrix over years

GDP per capita (current US\$) in 2013

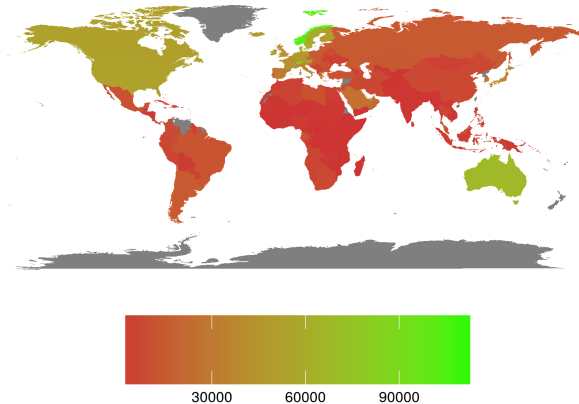


Рис.: GDP per capita, 2013

Population ages 65 and above (% of total) in 2010

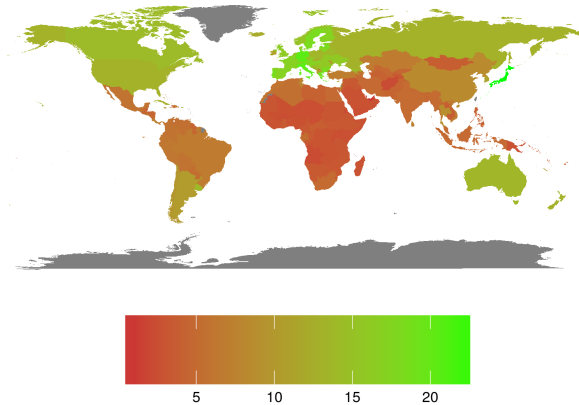


Рис.: Population aged over 65

## Unemployment Rates (2013)

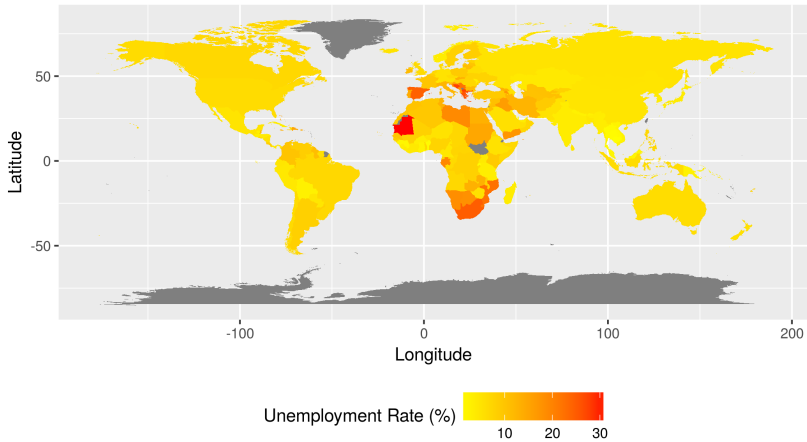


Рис.: World unemployment