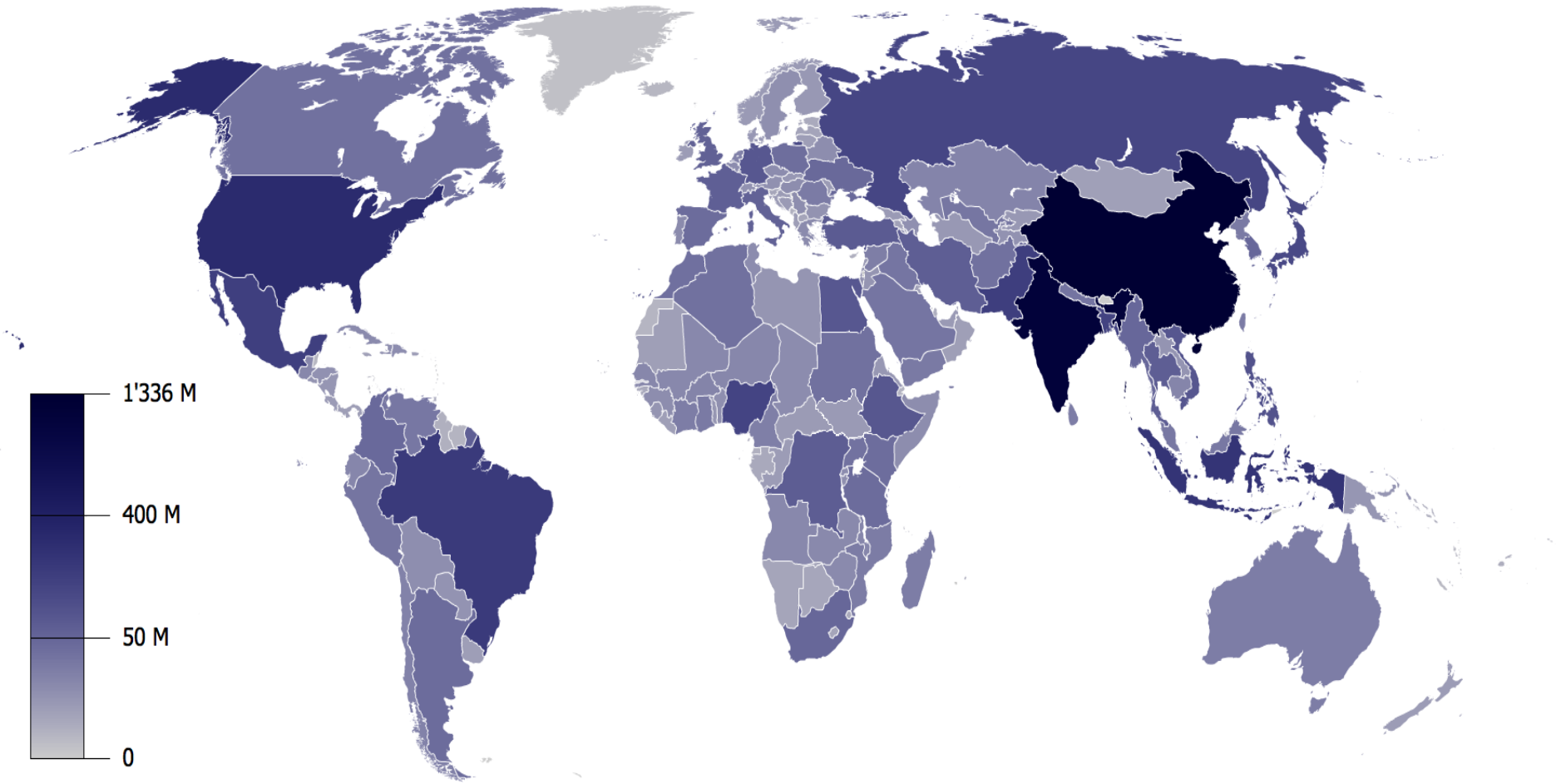
Weighting data in cross-national perspective

Marcin W. Zieliński
Polish Academy of Science
University of Warsaw


1'336 M
400 M
50 M
0

Population size weight

1. Calculate ratio of the size of country population and size of the country sample

$$WF_{pop} = \frac{N(population)}{n(sample)}$$

2. Scale for countries you want to combine:

$$sW_{pop} = \frac{WF_{pop}}{mean(WF_{pop})}$$

1st example

Data for 2006 ISSP:

	N (population)	n (sample)	Ratio (WF_{pop}) N/n	Scaled ratio *(*sWF_{pop}*)*
Poland	38.183.683	1263	30.232,528	30.232,528/13.932,523= **2,170**
Czech Rep.	10.384.603	1512	6.838,124	**0,491**
Slovakia	5.379.233	1138	4.726,918	**0,339**
			Mean(ratio)=13.932,523	Mean(scaled ratio)=1

1st example

Data for 2006 ISSP:

	N (population)	n (sample)	Ratio (WF_{pop}) N/n	Scaled ratio (sWF_{pop})
Poland	38.183.683	1263	30.232,528	30.232,528/13.932,523= **2,170**
Czech Rep.	10.384.603	1512	6.838,124	**0,491**
Slovakia	5.379.233	1138	4.726,918	**0,339**
			Mean(ratio)=13.932,523	Mean(scaled ratio)=1

RATIO ALWAYS HAS TO BE SCALED!

2nd example

- What if we want to compare European and North American countries?

2nd example

- What if we want to compare European and North American countries?

1. Calculate RATIO for all countries
2. Scale RATIO separately for European and North American countries
3. Weight your data by constructed weight

What changes population size weight

- It emphasizes the relative impact of the size of the population of a country on the outcome
- If the size of the country is related to the outcome then pop weight has bigger effect on the outcome

When pop weight really matters:

Dependent: life satisfaction

People from big countries tend to be more satisfied than people from small countries

When pop weight really matters:

Dependent: life satisfaction

People from big countries are as much satisfied as people from small countries

Is only population weighting sufficient?

- Pop weighting corrects only for size of the population and its impact on the outcome

$$WF_{gen} = WF_{des} * WF_{post} * WF_{pop} * \ldots$$

Differences in weight construction in cross-national comparisons

ISJP 1991:

USA=HH size, region, gender, age

Poland=HH size

Hungary=gender, employment status

Are the data comparable?

What to do?

- Calculate new weights using external sources of information (like e.g. UN data)

Advantages:

- The same factors taken into account
- Avoiding mistakes

Disadvantages:

- Practically weighting only by age and gender
- Loosing information about design weight factor
- Loosing information about other factors

What to do?

An example:

AFB, 2000, Republic of South Africa:

weight components:

province, race, gender, residental area, language (among Whites), housing type (among Blacks)

What to do?

- Do we know better than the authors of the study what factors should be included? Leave as it is?

Advantages:

  - Taking into account authors perspective
  - Design component where it was available

Disadvantages:

  - Comparability problem