The European Commission’s science and knowledge service

Joint Research Centre

European Commission

Step 4: Normalisation for Composite Indicators

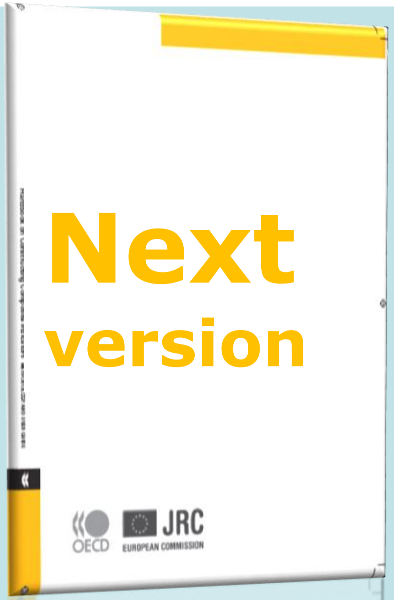
Maria Del Sorbo

COIN 2018 - 16th JRC Annual Training on Composite Indicators & Scoreboards
05-07/11/2018, Ispra (IT)

European Commission

Ten Steps Guide

for Composite Indicators & Scoreboards



- *Next version (2019-2020)*

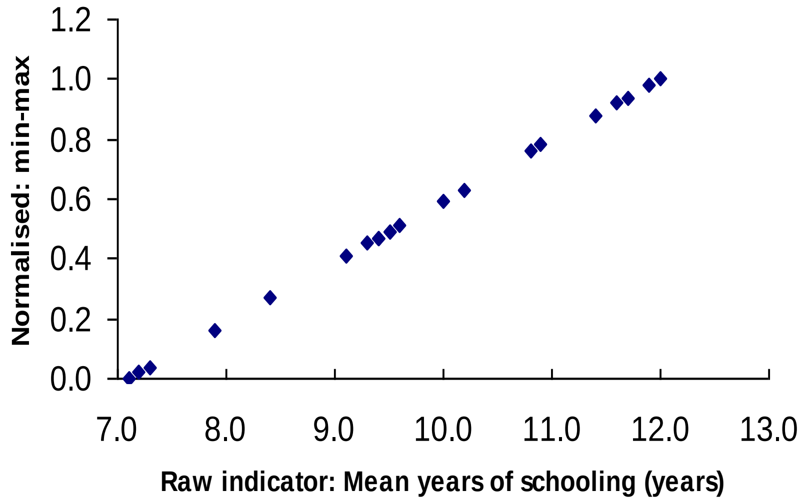
- Step 10. Visualisation & Communication
- Step 9. Back to the data
- Step 8. Robustness & Sensitivity
- Step 7. Statistical coherence
- Step 6. Aggregation
- Step 5. Weighting
- Step 4. Normalisation
- Step 3. Data treatment
- Step 2. Selection of indicators
- Step 1. Developing the framework

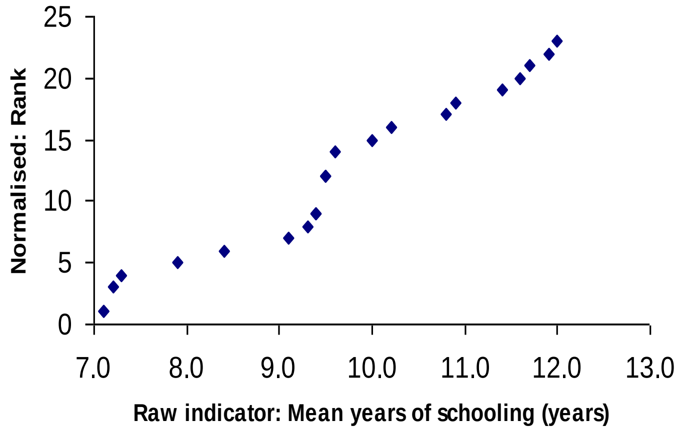
European Commission

Step 4

Normalisation for C.I.

- ✓ Make directional adjustments (so that higher scores correspond to better performance in all indicators or vice versa)
- ✓ Select a suitable normalisation method that respects the conceptual framework and the data properties





European Commission

Outline

- What is normalisation?
- Prior to normalisation
- Normalisation methods: min-max, distance to a reference, standardisation (or z-score), categorical scale, ranking and quantile empirical distribution
- Summary table
- Key messages

European Commission

What is normalisation?

Adjustments of distribution and scale of variables

It allows for aggregation of variables by averages (i.e. composite indicators) and for comparisons and robust aggregations

- different units of measurement **=> common scale**
- different ranges of variation **=> similar range of variation**

Different normalisation methods lead to different results
Choice based on the general objective

Normalising by…

Sample statistics

To obtain comparability across variables adjusting for:

1. different nature of indicators (positive vs. negative orientation towards the index)
2. different units of measurement across indicators
3. different ranges of variation

A common variable

To obtain comparability across countries

Dividing the raw data by size of population, land area, gross domestic product (GDP), or other **denominator** to make data comparable across countries

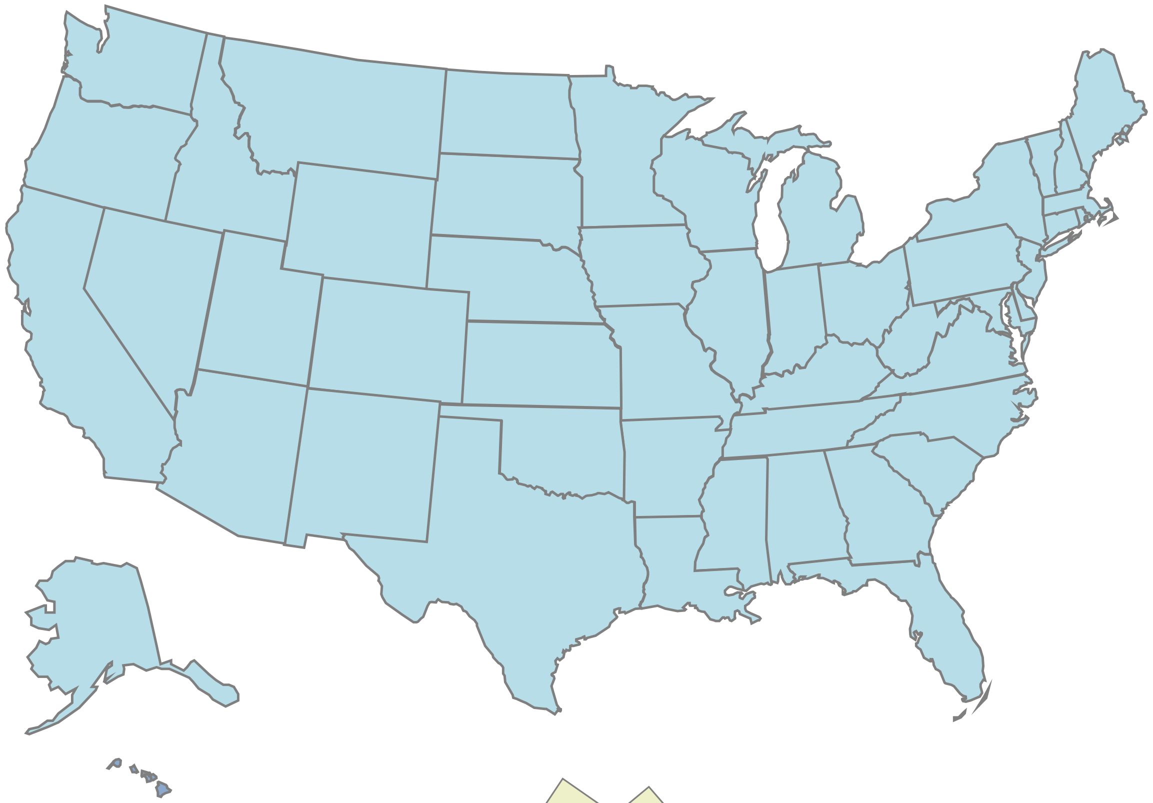
Prior to normalisation

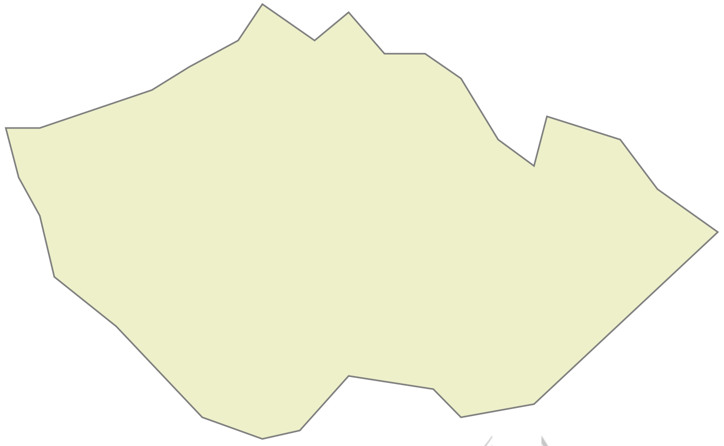
Comparisons across countries

Q: What to consider?

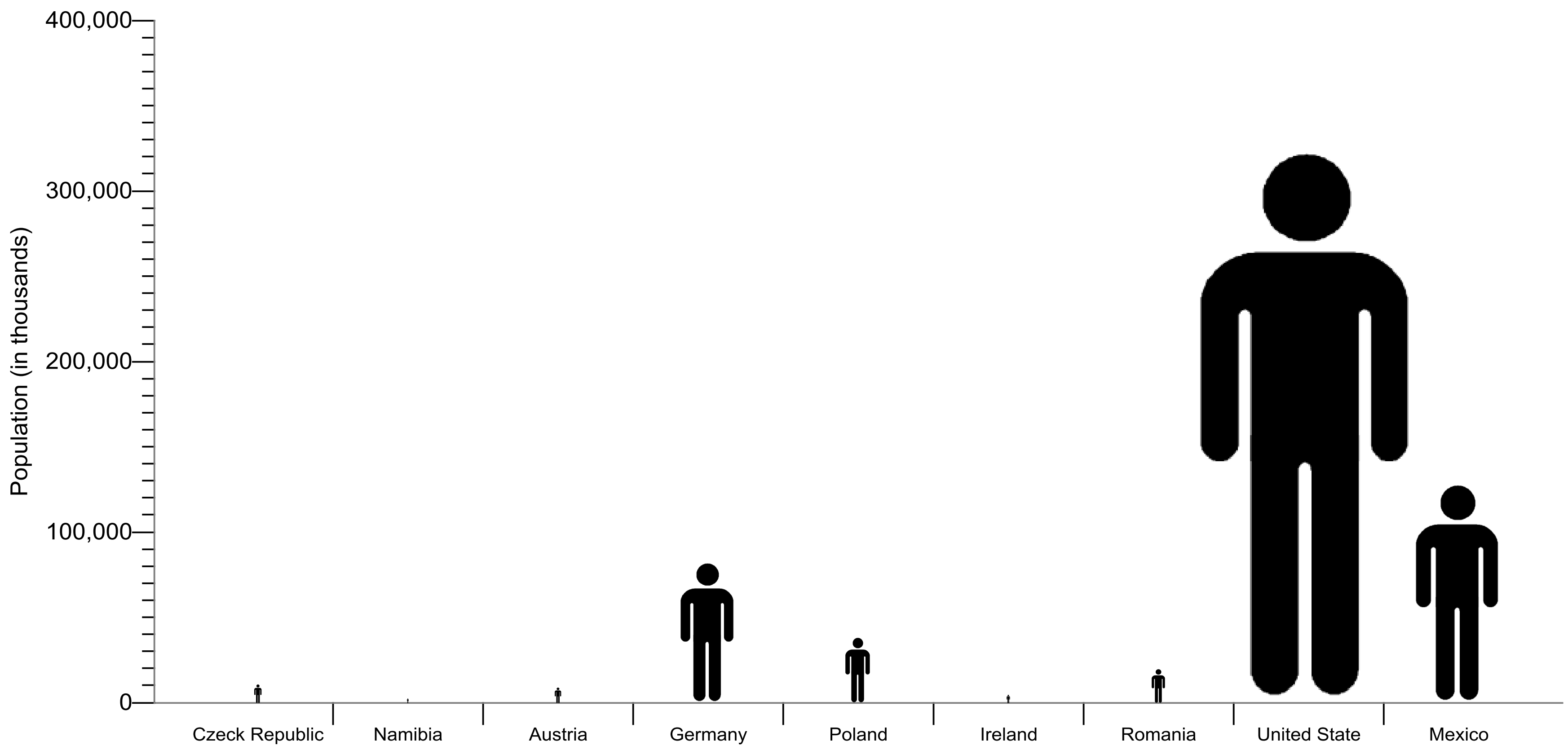
US, 322 Million inhabitants

Czech Republic, 10 Million inhabitants





World Population in 2016 (thousand)



	Czeck Republic	Namibia	Austria	Germany	Poland	Ireland	Romania	United State	Mexico
Population	10,610	2,479	8,712	81,914	38,224	4,726	19,778	322,179	127,540

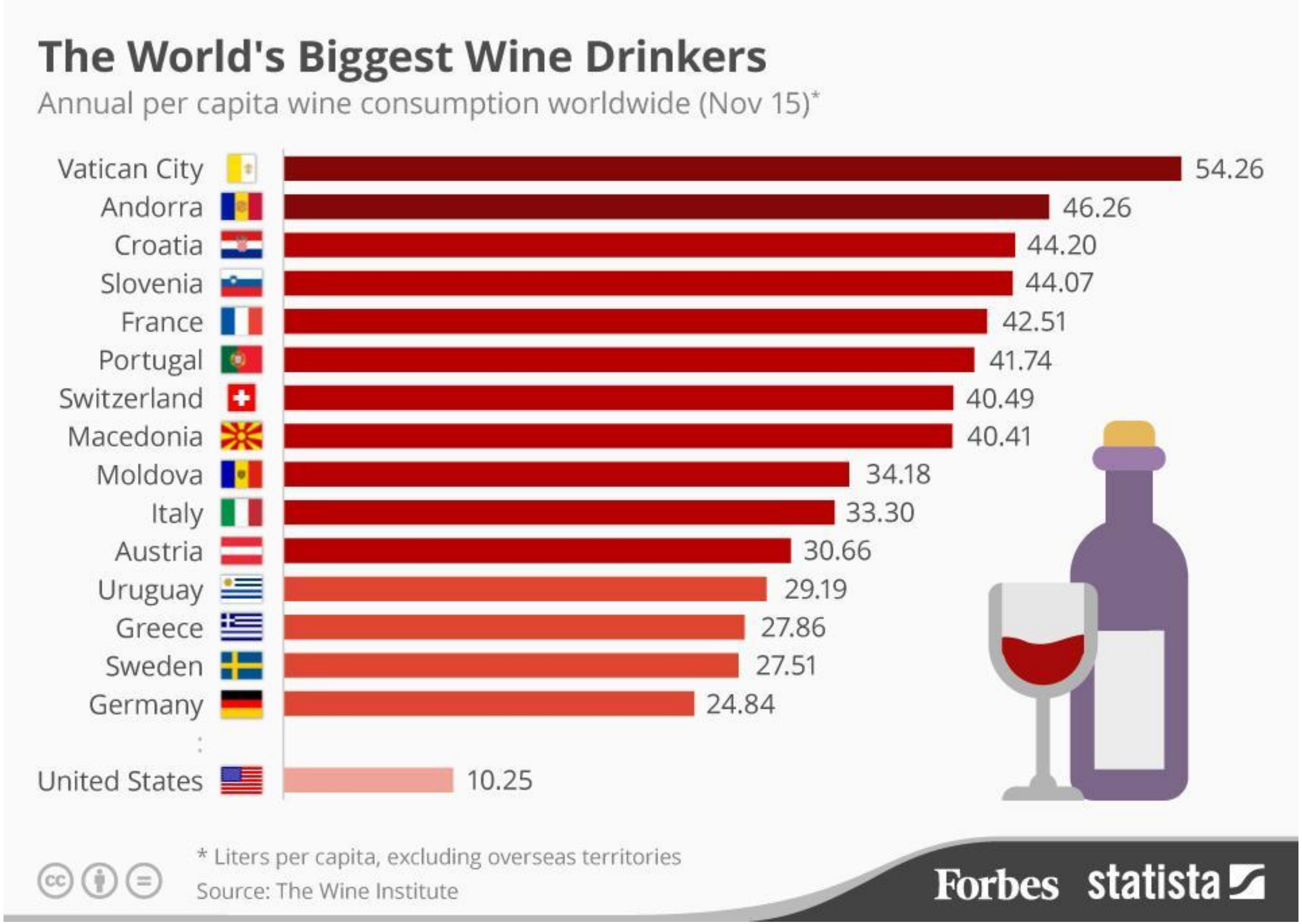
Source: United Nations, 2016

European Commission

Beer consumption - liters per capita	Czech Republic	Namibia	Austria	Germany	Poland	Ireland	Romania	US	Mexico
2016	143.3	108.0	106.0	104.0	100.8	98.2	94.1	74.8	65.1

World Beer Consumption per capita

https://www.kirinholdings.co.jp/english/news/2017/1221_01.html

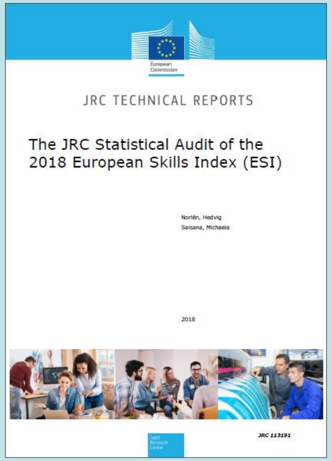

The World's Biggest Wine Drinkers
Annual per capita wine consumption worldwide (Nov 15)*
Vatican City 54.26
Andorra 46.26
Croatia 44.20
Slovenia 44.07
France 42.51
Portugal 41.74
Switzerland 40.49
Macedonia 40.41
Moldova 34.18
Italy 33.30
Austria 30.66
Uruguay 29.19
Greece 27.86
Sweden 27.51
Germany 24.84
United States 10.25
* Liters per capita, excluding overseas territories
Source: The Wine Institute
Forbes statista

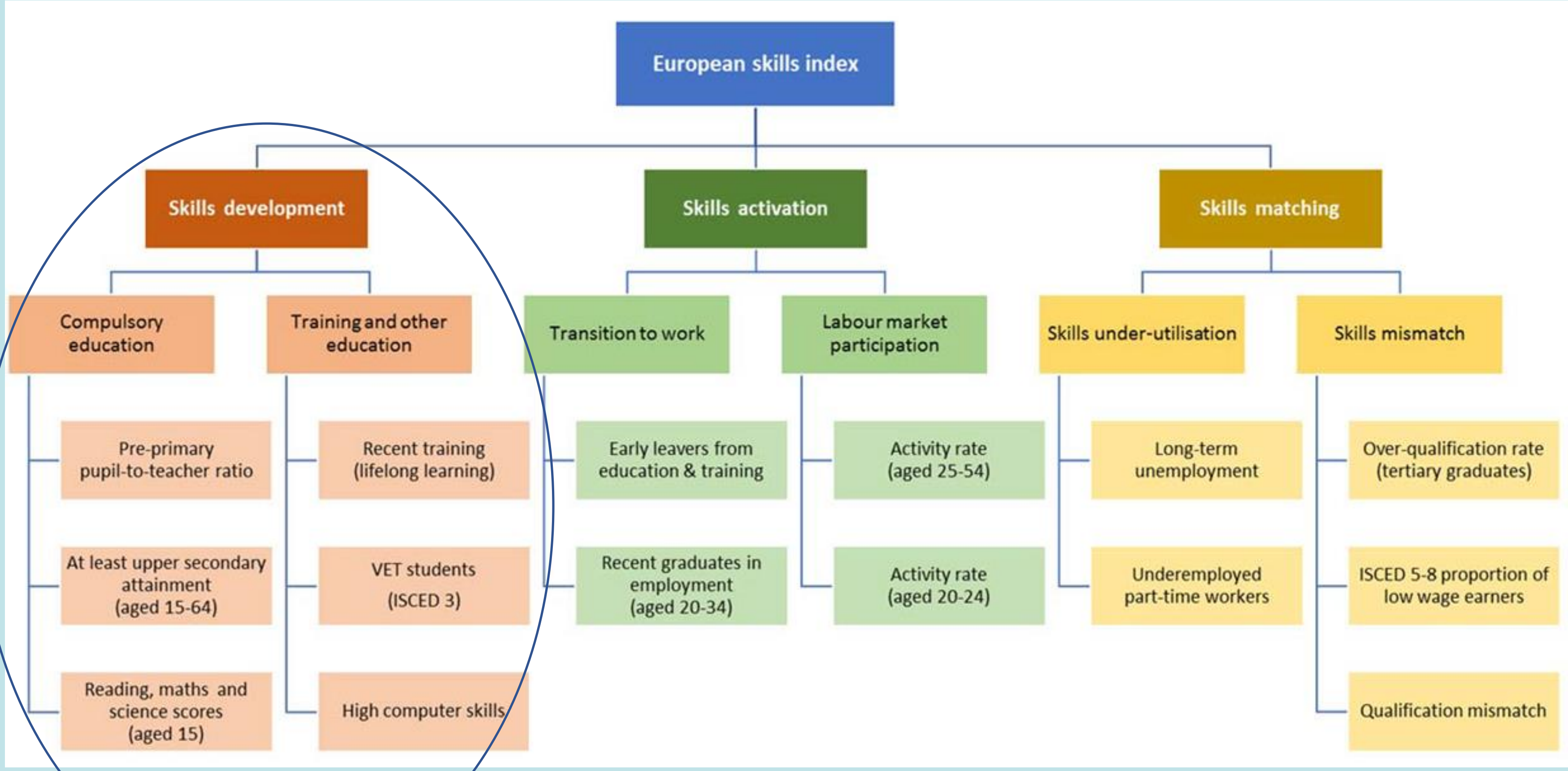
Data normalisation

- It is about changing the scale of the data
- Common practice: scaling data between [0, 1] or [0, 100]
- Mixed scales (e.g., European skills Index: PISA scores and unemployment rate have different ranges)

European Commission

Example: European Skills Index





Source: European Skills Index (2018), Cedefop

Range of variation
Example: European Skills Index

Table 4.2: Summary statistics

Indicator (unit)	Range
Pre-primary pupil-to-teacher ratio (students per teacher)	[6.4, 21.5]
Share of population aged 15-64 with at least upper secondary education (%)	[47.1, 87.6]
Reading, maths & science scores (aged 15) (PISA score)	[437.7, 524.3]
Recent training (%)	[1.2, 29.6]
VET students (%)	[1.2, 73.2]
High computer skills (%)	[7.0, 46.0]

European Commission

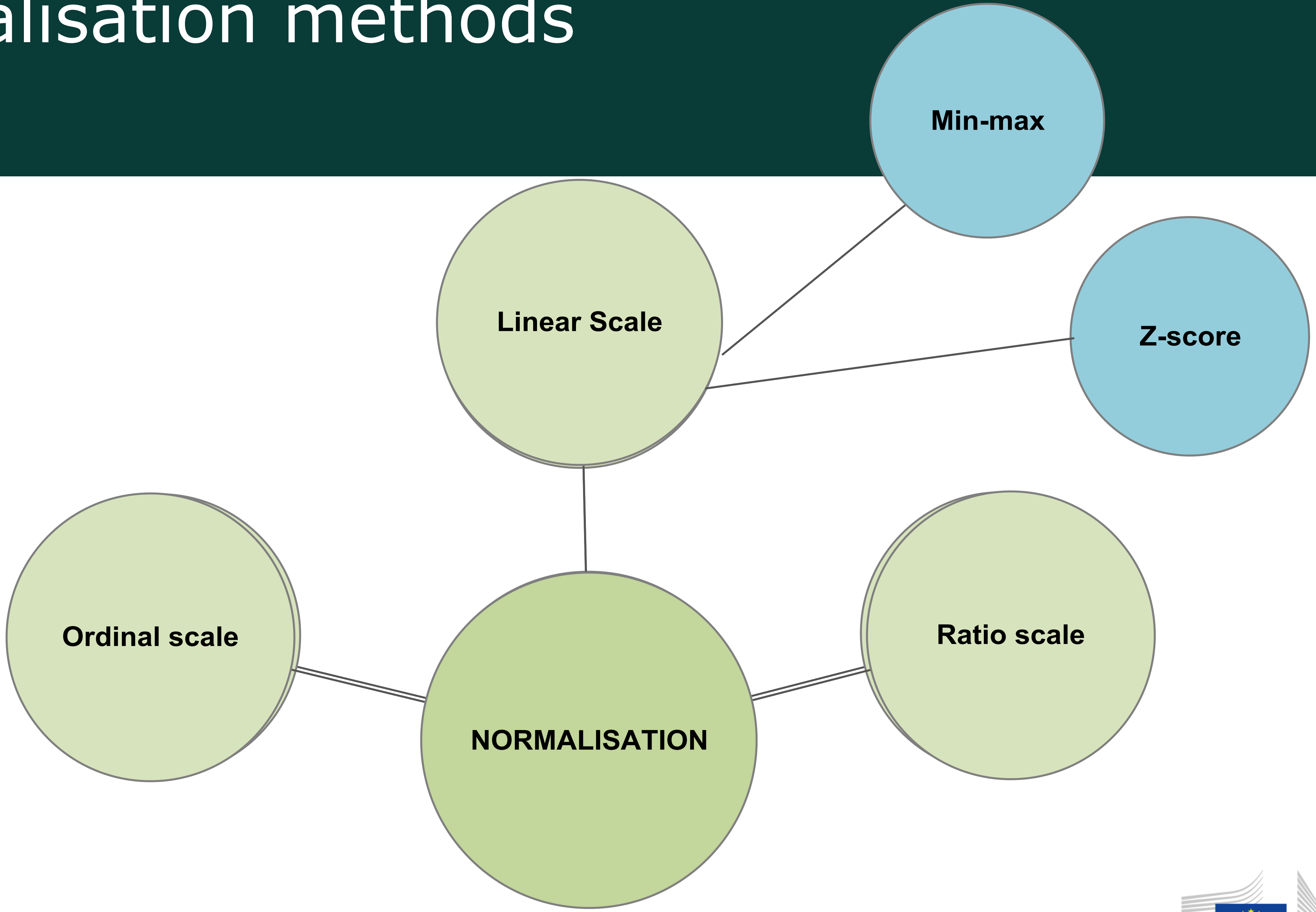
The W. EDWARDS Deming Institute

W. Edwards Deming

If you change the rule for counting people, you come up with a new number.

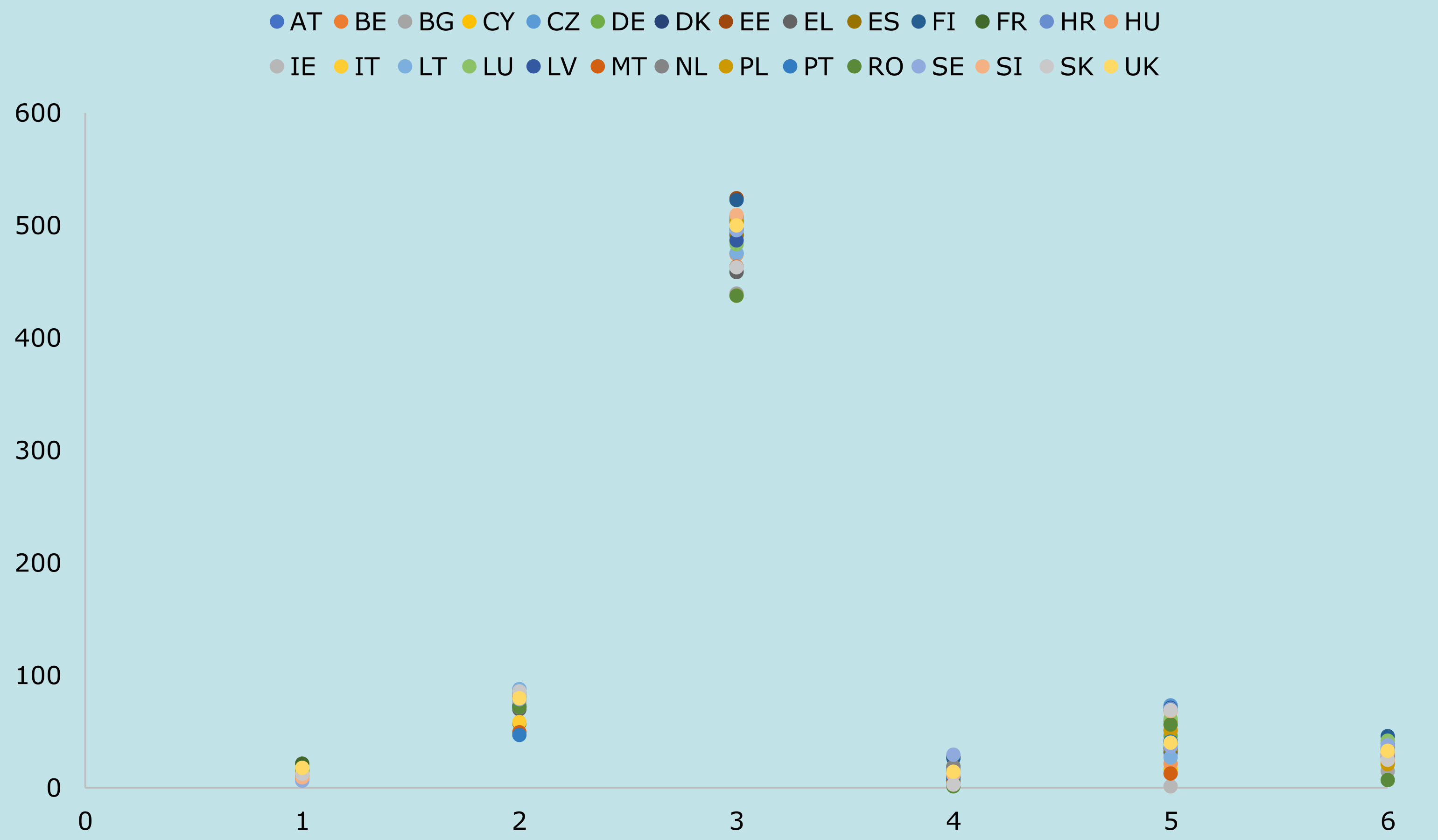
source: quotes.deming.org/10217

Normalisation methods



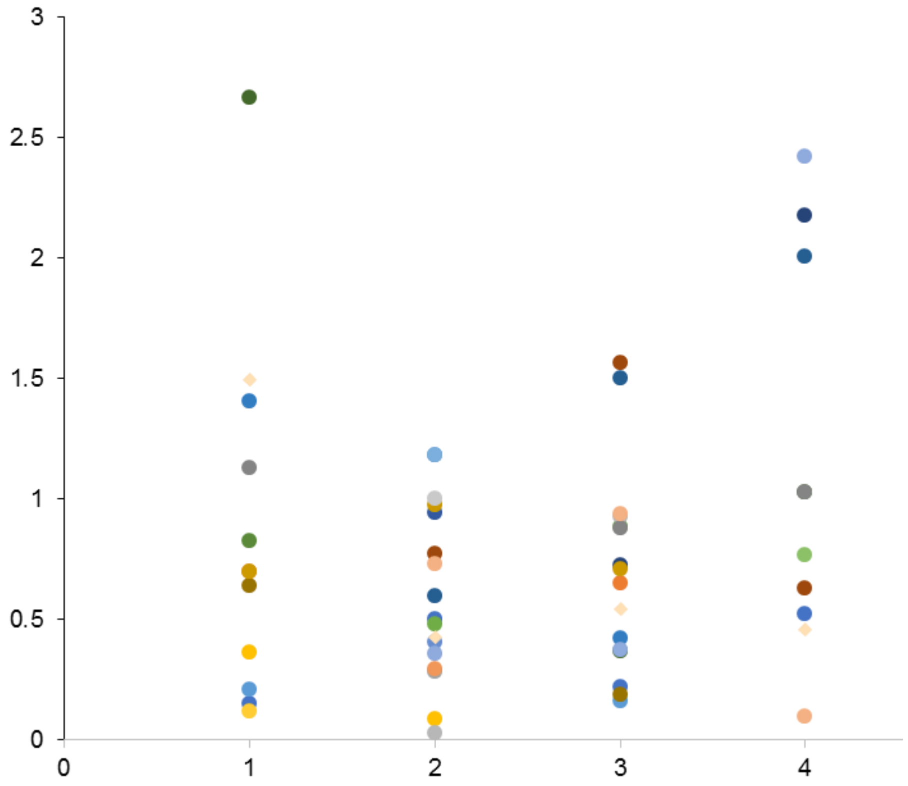
European Commission

Data before normalization

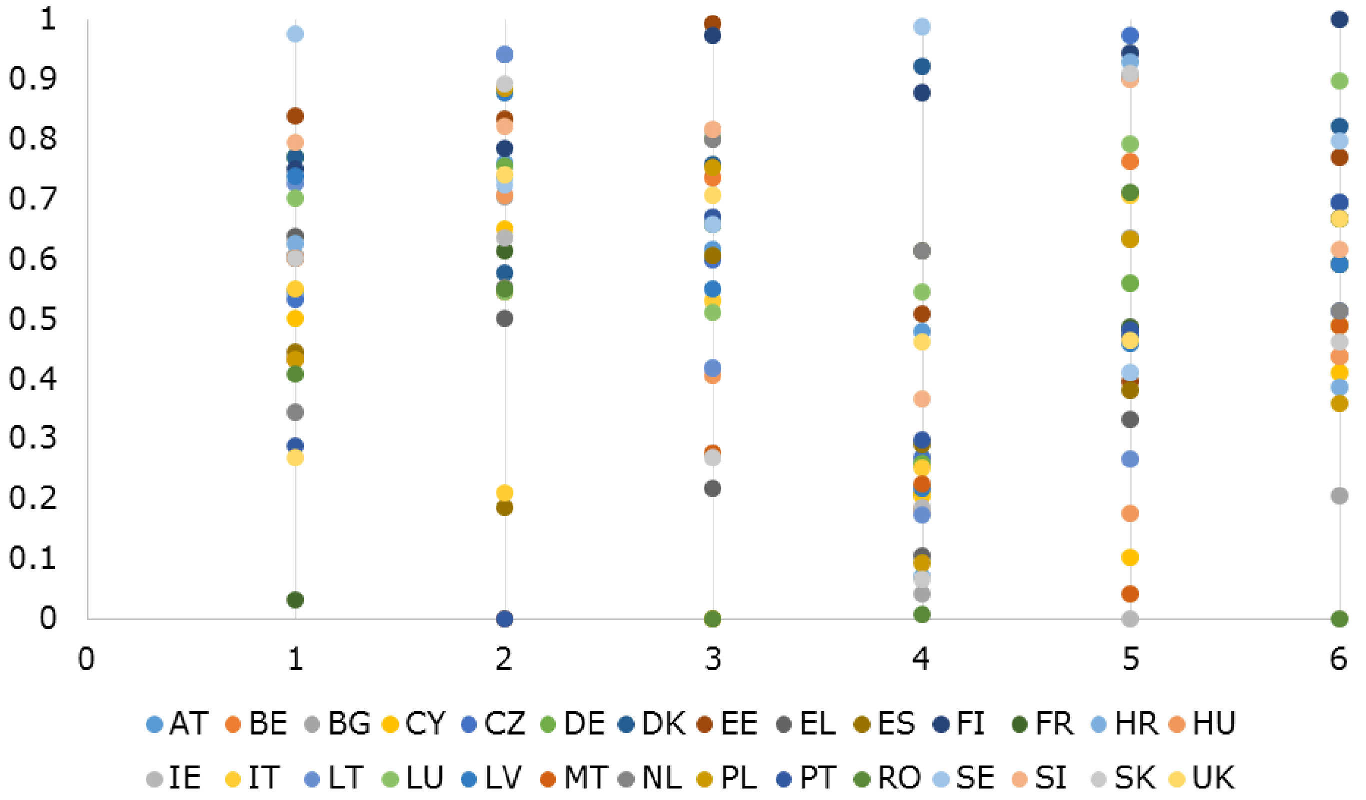


European Commission

Z-score



Min-max



Normalisation methods – linear scale
Standardisation (Z-score)

$$I = \frac{x - \bar{x}}{s} \quad \sim \quad z = \frac{x - \mu}{\sigma}$$

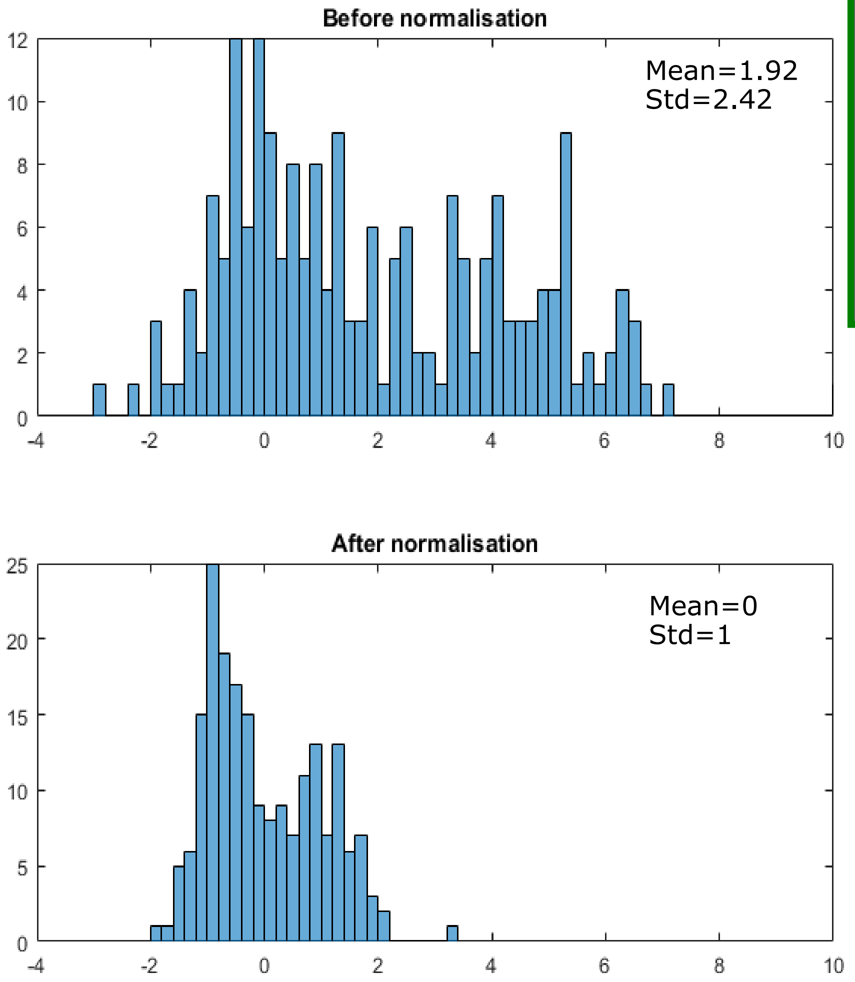
- Imposes a distribution with mean zero and variance 1
- **Normalised indicators have:**

✓ same variance (=1)

✓ not necessarily same range of variation

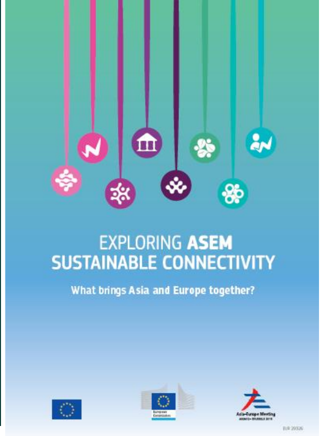
- **Method**: sensitive to extreme values/outliers

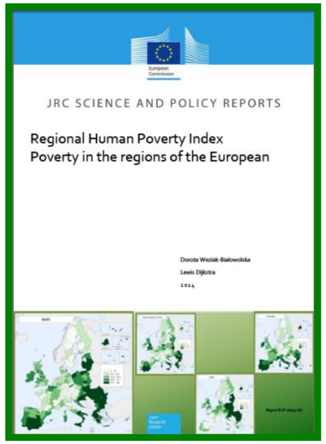
Standardised scores which are below average are negative => implications on the use of geometric average as an aggregation method



European Commission

Normalisation methods – linear scale, Min-max





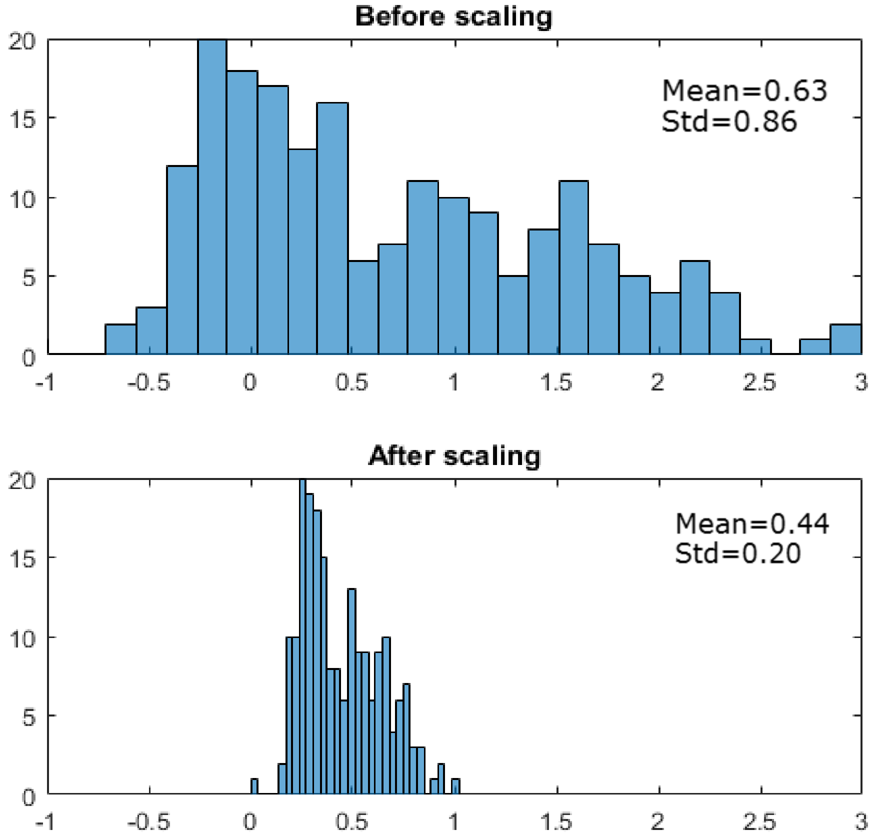
$$I = \frac{x - \min}{\max(x) - \min(x)}$$

Normalised indicators have:

- ✓ same range of variation e.g., [0,1]
- ✓ not same variance

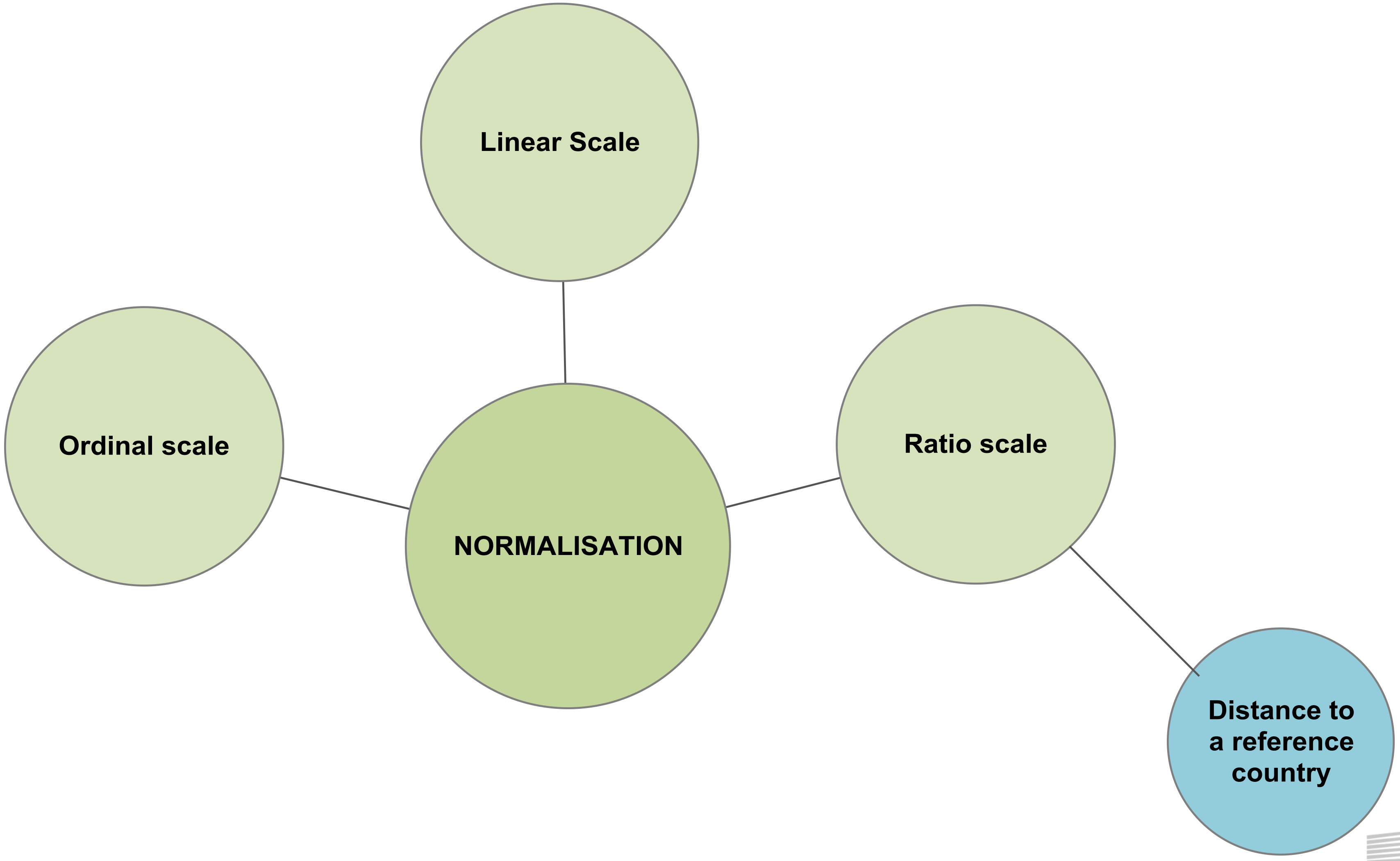
Method: sensitive to extreme values/outliers

Rescaling is easier to communicate to a wider public because it normalises indicators to an identical range [0, 1], [0, 100], where usually higher scores represent better achievement.



European Commission

Normalisation methods: Ratio scale



Normalisation methods: Ratio scale - Distance to a reference country

The reference country can be:

- the group leader or an external benchmark country
- a hypothetical country (target to be reached in a given timeframe)
- or an aggregate/average (eg., EU28, world)

The reference value could be fixed at a specific point in time (e.g. initial t_0) to account for the indicator evolution across time

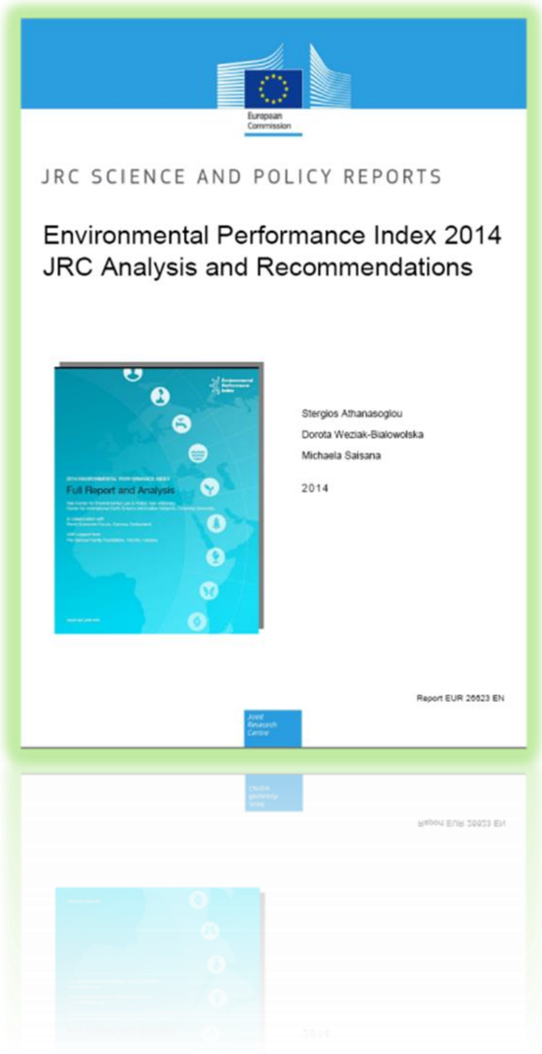
Normalised indicators do not have:

✓ same range
✓ same variance

Method: sensitive to extreme values/outliers

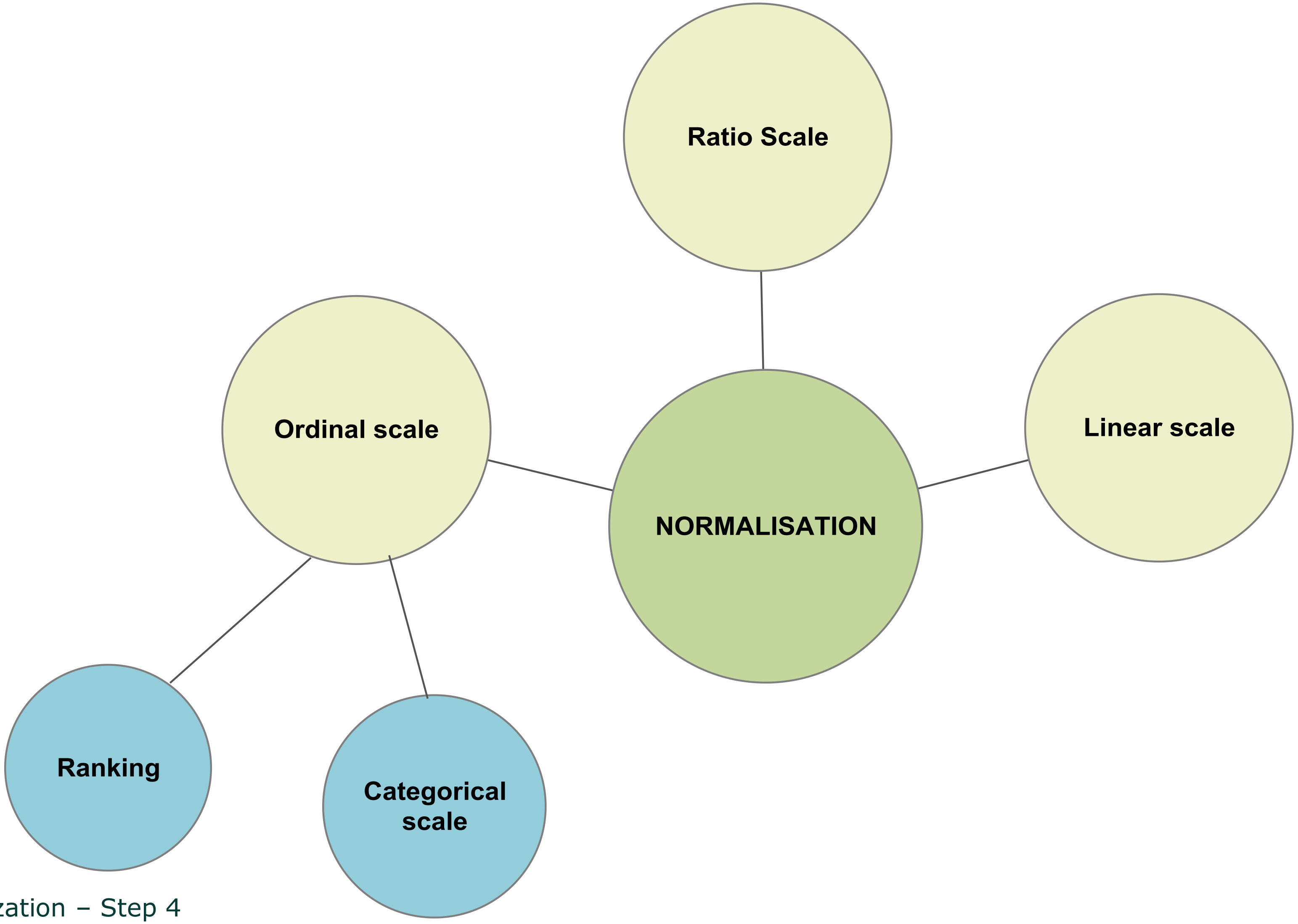
$$\mathrm{I}_c = \frac{x_c}{x_{\bar{c}}}$$

$$\mathrm{I}_c = \frac{x_c^{t}}{x_c^{t_0}}$$



European Commission

Ordinal scale: Ranking across countries and categorical scale



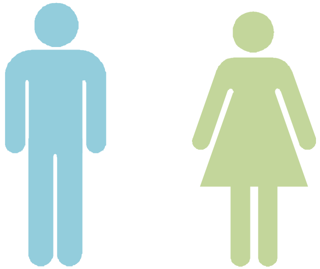
European Commission

Normalisation methods – Ordinal scale

Categorical scales

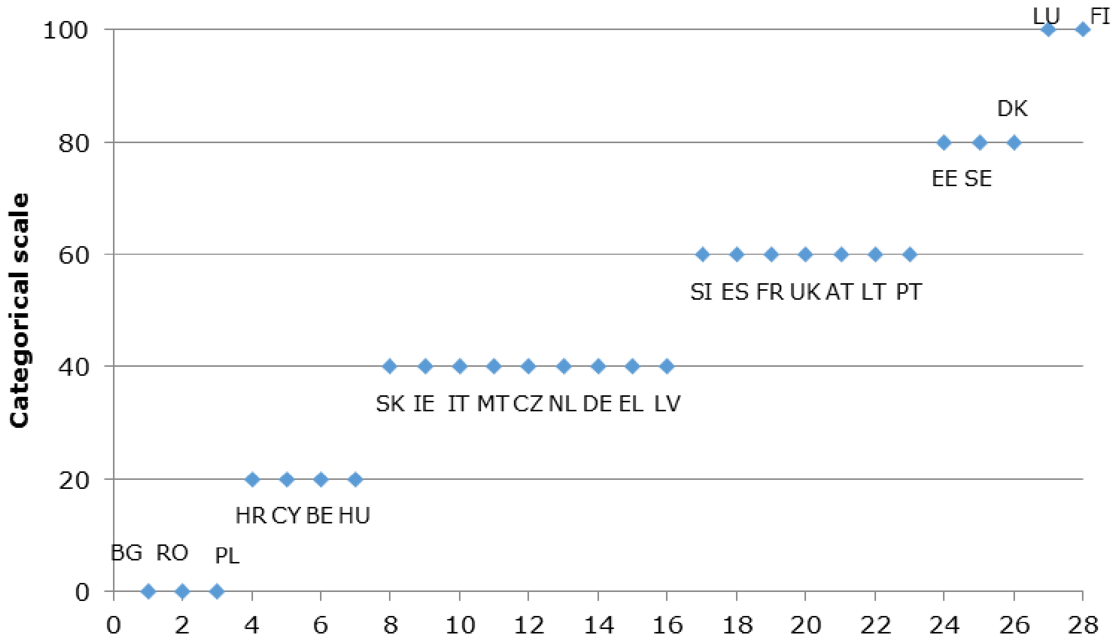
- The score of each indicator is based on categories (nominal or numerical)
- Each category represents a portion of the range of the variable
- Each category score can be based on the percentile of the distribution of the indicator across countries
- Remember to justify the choice of intervals and scores
- **Normalised indicators have**:
  - Same range [0, 100]
  - Same variance: if there are NOT tied ranks
- **Method**: not sensitive to extreme values/outliers
- Distribution: NO uniform

Nominal scales (e.g. gender, blood type: A, B, AB, O)



Numerical scales

$$I_{q,c}^{t} = \begin{cases} 0 & if \;\; p^{0} \le x_{q,c}^{t} < p^{10} \\ 20 & if \;\; p^{10} \le x_{q,c}^{t} < p^{25} \\ 40 & if \;\; p^{25} \le x_{q,c}^{t} < p^{50} \\ 60 & if \;\; p^{50} \le x_{q,c}^{t} < p^{75} \\ 80 & if \;\; p^{75} \le x_{q,c}^{t} < p^{90} \\ 100 & if \;\; p^{90} \le x_{q,c}^{t} \le p^{100} \end{cases}$$



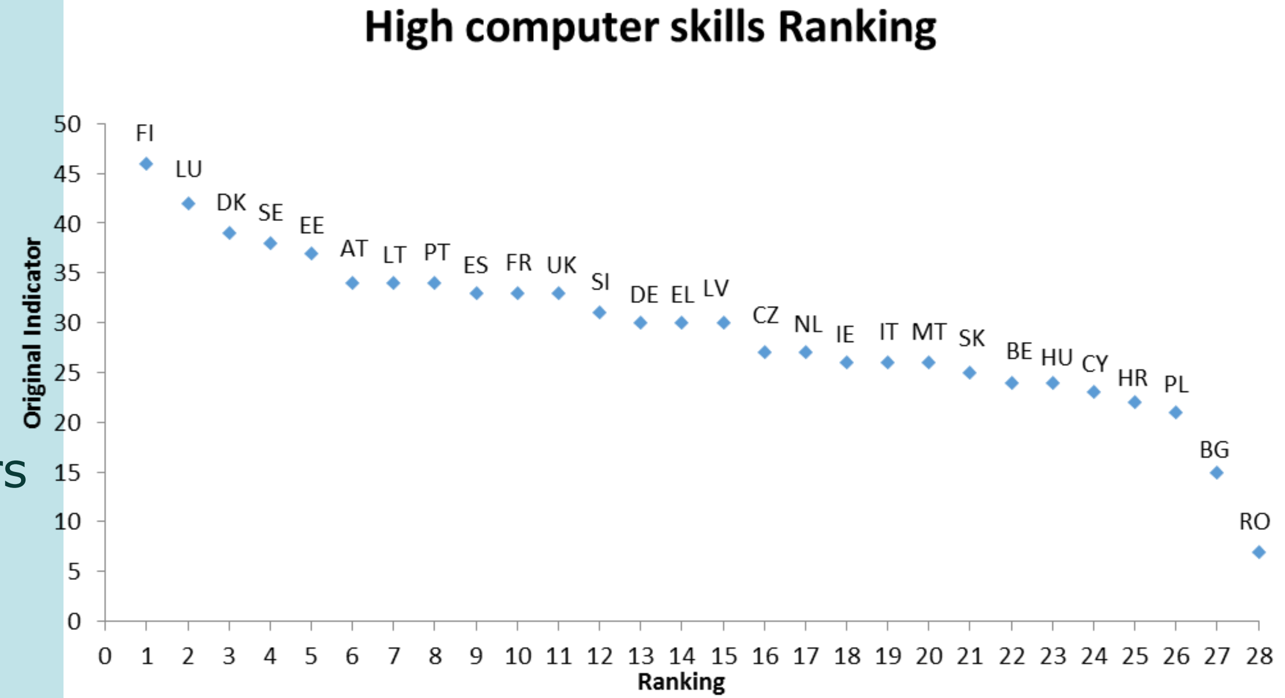
European Commission

Normalisation methods – Ordinal scale

Ranking across countries

$$I = rank(x)$$

- Scores are replaced by ranks – e.g. the highest score receives the first ranking position (rank 1)
- Uses ordinal information only – information on levels is not maintained
- **Normalised indicators have**:
  - ✓ the same range: [1,n] (n = no. of countries)
  - ✓ the same variance 67.67 (n=28)
- **Method**: not sensitive to extreme values/outliers
- Distribution: Uniform



Source: European Skills Index, 2018

Normalisation methods: Quantile empirical distribution

Definition: quantile normalization is a technique for making two distributions identical in statistical properties

$$u = \frac{rank(x)}{N+1}$$

- where N is the sample size;
- rank(x) is the rank associated to each realization

The quantile normalisation method allow to gather a variable range [0, 1]

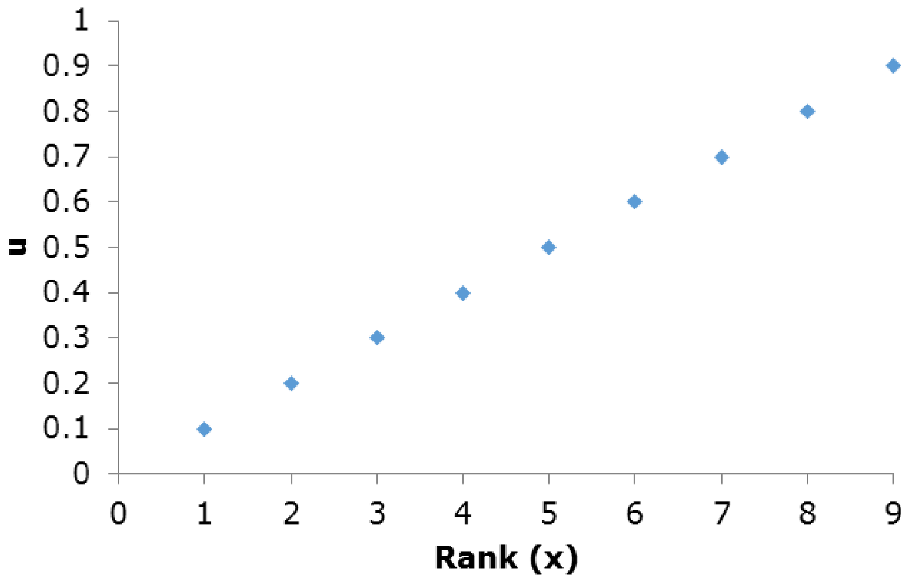
Normalisation methods: Quantile empirical distribution

For example: if x = (10, 2, 8, 81, 38, 4, 19, 322, 127)

Then, rank (4, 1, 3, 7, 6, 2, 5, 9, 8); u = (0.4, 0.1, 0.3, 0.7, 0.6, 0.2, 0.5, 0.9, 0.8) giving to 322 the 9th rank

$$u = \frac{rank(x)}{N+1}$$

Example: let's consider the population in million by country – slide 9



x= million inhabitants	Rank(x)	u
2	1	0.1
4	2	0.2
8	3	0.3
10	4	0.4
19	5	0.5
38	6	0.6
81	7	0.7
127	8	0.8
322	9	0.9

European Commission

Summary table

Adjustments/Method	Quantile empirical distribution	Ranking	Categorical scale	Z-score	Min-max	Distance to a reference country
Unit of measurement	Y	Y	Y	Y	Y	Y
Variance	Y	Y	Y/N	Y	N	N
Range of variation	Y	Y	Y	N	Y	N
Extreme values*	Y	Y	Y	N	N	N
Distribution*	Y	Y	Y/N	N	N	N
Extreme values*	non-sensitive to extreme values					
Distribution*	the distribution will be the same for the normalised indicators					

Yes is valid only if there are not tied ranks

European Commission

Key messages

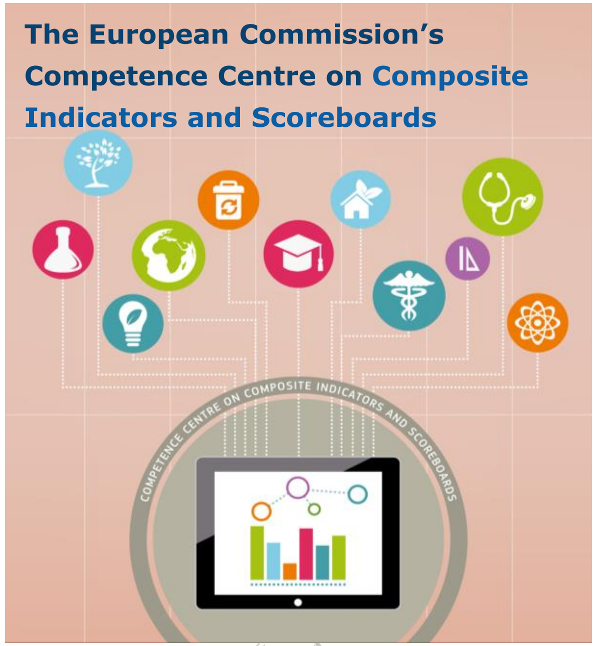
- There are different normalisation methods leading to different results
- Their selection depends on the data structure and what we want to achieve: e.g., **same variance=> Z-score**; **same range of variation =>min-max.** It is crucial to understand the properties of each
- Alternative normalisation methods can be included as an additional dimension in the uncertainty/sensitivity analysis
- Normalisation is the step the lowest impact on the ranking (based on the majority of the case studies)

THANK YOU

Welcome to email us at: jrc-coin@ec.europa.eu

COIN in the EU Science Hub
https://ec.europa.eu/jrc/en/coin

COIN tools are available at:
https://composite-indicators.jrc.ec.europa.eu/



European Commission

References

- Becker, W., M. Saisana, P. Paruolo, and I. Vandecasteele. 2017. 'Weights and Importance in Composite Indicators: Closing the Gap. Ecological Indicators 80: 12–22.
- Bichindaritz I., Big Data, Genes, and Medicine, the State University of New York (SUNY), https://www.coursera.org/lecture/data-genes-medicine/data-normalization-jGN7k
- Cohen, A., Saisana, M., 2014, Quantifying the Qualitative: Eliciting Expert Input to Develop the Multidimensional Poverty Assessment Tool, Journal of Development Studies 50(1): 35-50.
- European Union, European Centre for the Development of Vocational Training (Cedefop). 2018, European Skills Index technical report. Thessaloniki, Greece.
- Forbes-Statista, 2016, The world's biggest wine drinkers, https://www.forbes.com/sites/niallmccarthy/2016/05/17/the-worlds-biggest-wine-drinkers-infographic/#2c1468a77a57
- Kirin Holdings Company. Kirin Beer University Report Global Beer Consumption by Country in 2016, Global Beer Consumption by Country in 2016, Table 1 Per Capita Beer Consumption by Country in 2016, Table 3, https://www.kirinholdings.co.jp/english/news/2017/1221_01.html
- OECD/JRC, 2008, *Handbook on Constructing Composite Indicators. Methodology and user Guide*, OECD Publishing, ISBN 978-92-64-04345-9.

European Commission

References

- Organisation Internationale de la Vigne ed du Vin, 2016, OIV, State of the viticulture World Market, p. 10.
- Paruolo P., Saisana M., Saltelli A., 2013, Ratings and Rankings: voodoo or science?. J Royal Statistical Society A 176(3), 609-634.
- Saisana, M., and Saltelli, A., 2011, Rankings and Ratings: Instructions for use, Hague Journal on the Rule of Law 3(2), 247-268.
- Saisana M., D’Hombres B., Saltelli A., 2011, Rickety Numbers: Volatility of university rankings and policy implications. Research Policy 40, 165–177.

European Commission

Technical Appendix

Global Beer Consumption by Country in 2016

2016 Ranking	2015 Ranking	Country	2016			2015	
			Total Consumption (thousand kl)	Global Market Share	Growth Rate 2015-2016	Total Consumption (thousand kl)	Global Market Share
1	1	China	41,772	22.4%	-3.4%	43,264	23.0%
2	2	United States	24,245	13.0%	0.6%	24,106	12.8%
3	3	Brazil	12,654	6.8%	-2.7%	13,008	6.9%
4	5	Germany	8,412	4.5%	-0.5%	8,450	4.5%
5	4	Russia	8,405	4.5%	-1.8%	8,559	4.6%
6	6	Mexico	7,988	4.3%	8.4%	7,371	3.9%
7	7	Japan	5,251	2.8%	-2.4%	5,380	2.9%
8	8	United Kingdom	4,373	2.3%	-0.9%	4,413	2.3%
9	9	Vietnam	4,117	2.2%	7.4%	3,832	2.0%
10	11	Spain	3,909	2.1%	2.3%	3,821	2.0%
11	10	Poland	3,892	2.1%	1.8%	3,823	2.0%
12	12	South Africa	3,145	1.7%	2.4%	3,072	1.6%
13	13	India	2,701	1.4%	9.9%	2,457	1.3%
14	14	Colombia	2,357	1.3%	3.0%	2,289	1.2%
15	15	South Korea	2,160	1.2%	1.0%	2,139	1.1%
16	17	Canada	2,093	1.1%	-0.3%	2,100	1.1%
17	18	France	2,061	1.1%	2.4%	2,012	1.1%
18	19	Czech Republic	1,959	1.0%	1.5%	1,930	1.0%
19	20	Thailand	1,910	1.0%	1.5%	1,881	1.0%
20	23	Romania	1,826	1.0%	1.6%	1,797	1.0%
21	22	Argentina	1,778	1.0%	-5.2%	1,875	1.0%
22	25	Italy	1,745	0.9%	0.8%	1,730	0.9%
23	21	Ukraine	1,743	0.9%	-7.2%	1,878	1.0%
24	24	Australia	1,735	0.9%	0.1%	1,735	0.9%
25	27	Philippines	1,620	0.9%	6.0%	1,528	0.8%

Source: Kirin Company, Limited, (2016)

Global Wine Consumption

Main wine-consuming countries[a]

mhl	2012	2013	2014	2015[b]	2016[c]	2016/2015 Variation in volume	2016/2015 Variation in %
United States	30,0	30,2	30,4	31,0	31,8	0,8	2,5%
France	28,0	27,8	27,5	27,2	27,0	-0,2	-0,7%
Italy	21,6	20,8	19,5	21,4	22,5	1,1	5,3%
Germany	20,3	20,4	20,2	20,6	20,2	-0,4	-1,8 %
China	17,1	16,5	15,5	16,2	17,3	1,1 0,2	6,9%
United Kingdom *	12,8	12,7	12,6	12,7	12,9	0,0	1,4%
Spain	9,9	9,8	9,9	10,0	9,9		-0,4%
Argentina	10,1	10,4	9,9	10,3	9,4	-0,9	-8,3 %
Russia	11,3	10,4	9,6	9,3	9,3	0,0 0,1	0,3%
Australia	5,4	5,4	5,4	5,3	5,4	0,1	2,4%
Canada	4,9	4,9	4,7	4,9	5,0	0,0	3,1%
Portugal	5,0	4,8	4,7	4,8	4,8	0,1	0,1% 3,1%
South Africa	3,6	3,7	4,0	4,2	4,4		
Romania	4,3	4,6	4,7	3,9	3,8	-0,2	-4,5%
Japan	3,1	3,4	3,5	3,5	3,5	0,0	-0,3%
Netherlands	3,5	3,5	3,4	3,5	3,4	-0,1	-2,3%
Belgium	2,9	2,9	2,7	3,0	3,0	0,0	1,1%
Brazil	3,2	3,5	3,5	3,3	2,9	-0,4	-12,0%
Switzerland	2,7	2,7	2,8	2,9	2,8	-0,1	-1,8%
Austria	2,7	2,8	3,0	2,4	2,4	0,0	2,0%
Serbia	2,3	2,3	2,4	2,4	2,4	0,0	-0,9%
Sweden	2,3	2,4	2,3	2,4	2,3	-0,1	-3,3%
Greece	3,1	3,0	2,6	2,4	2,3	-0,1	-4,4%
Chile	3,2	2,9	3,0	2,1	2,2	0,1	4,8%
Hungary	2,0	1,9	2,2	2,2	1,9	-0,3	-12,7%
Denmark	1,5	1,6	1,6	1,6	1,6	0,0	0,0%
Croatia	1,4	1,4	1,2	1,1	1,2	0,1	6,7%
Poland	0,9	0,9	1,0	1,1	1,1	0,1	4,9%
Bulgaria	1,0	0,8	0,9	1,0	1,0	0,0	3,4%
World total	**244**	**243**	**240**	**241**	**242**	**0,9**	**0,4%**

Sources: OIV, OIV Experts, Trade Press

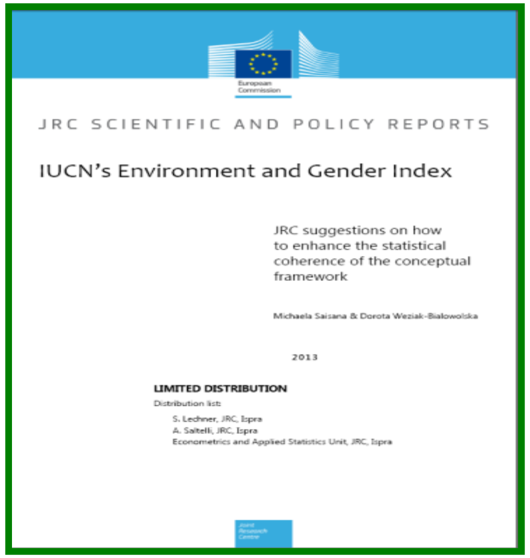
a) Countries for which information has been provided with wine consumption of more than 1 mhl
b) 2015: provisional data
c) 2016: forecasted data
* Apparent consumption calculated by "Production + Imports - Exports" data for 2015 and 2016

Source: Organisation Internationale de la Vigne ed du Vin, 2016

Min-Max

A simple alternative:

$$I_{q,c}^{t} = \frac{x_{q,c}^{t} - \min_{c,t}(x_q)}{\max_{c,t}(x_q) - \min_{c,t}(x_q)}$$



where the minimum and maximum for each indicator are calculated across countries and time. The normalized indicators have values between 0 and 1.

When **data for a new time point** become available the global minimum and/or the maximum may be affected. To maintain comparability between the existing and the new data, the composite indicator for the existing data must be re-calculated.

European Commission

Min-Max

The expression

$$\mathrm{I}^{\mathrm{t}}_{\mathrm{q,c}} = \frac{x^{t}_{q,c} - \min_{c}\left(x^{t_0}_{q}\right)}{\max_{c}\left(x^{t_0}_{q}\right) - \min_{c}\left(x^{t_0}_{q}\right)}$$

is sometimes used in time-dependent studies. However, if:

$$x^{t}_{q,c} > \max_{c}\left(x^{t_0}_{q}\right)$$

the normalised indicator would be larger than 1

Categorical scale: original score and rescaling

Country	High computer skills	Categorical Scale	th Percent
FI	46.00	100	100-90
LU	42.00	100	100-90
DK	39.00	80	75-90exl
SE	38.00	80	75-90exl
EE	37.00	80	75-90exl
AT	34.00	60	50-75exl
LT	34.00	60	50-75exl
PT	34.00	60	50-75exl
ES	33.00	60	50-75exl
FR	33.00	60	50-75exl
UK	33.00	60	50-75exl
SI	31.00	60	50-75exl
DE	30.00	40	25-50exl
EL	30.00	40	25-50exl
LV	30.00	40	25-50exl
CZ	27.00	40	25-50exl
NL	27.00	40	25-50exl
IE	26.00	40	25-50exl
IT	26.00	40	25-50exl
MT	26.00	40	25-50exl
SK	25.00	40	25-50exl
BE	24.00	20	10-25exl
HU	24.00	20	10-25exl
CY	23.00	20	10-25exl
HR	22.00	20	10-25exl
PL	21.00	0	0-10exl
BG	15.00	0	0-10exl
RO	7.00	0	0-10exl

European Commission