Assessing the statistical quality of Composite Indicators

The JRC methodology and the analysis of the ICT Development Index

Speaker: Giulio Caperna
Contributions by Eleni Papadimitriou
and JRC-COIN
ITU Building, Geneva

Joint
Research
Centre

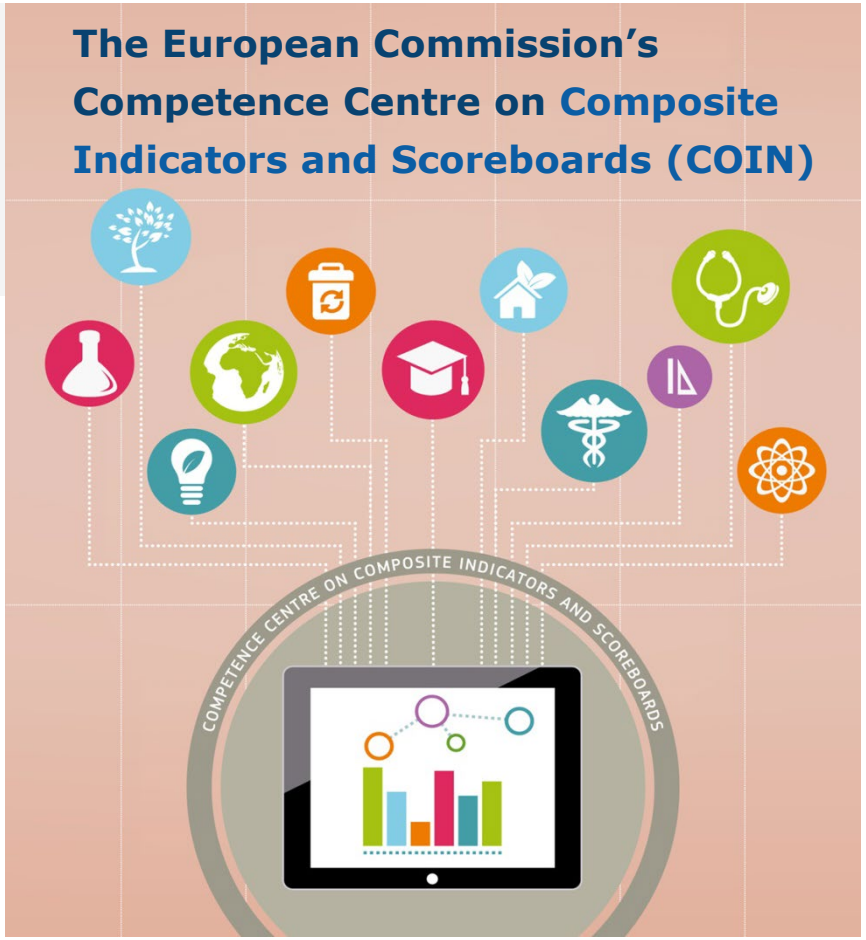
Pre-Audit of the IDI (v3)

- JRC-COIN has been invited by the (ITU) to pre-audit the new edition of the ICT Development Index (IDI)
- JRC-COIN's statistical audits provide statistical assessment of composite indicators, and contribute improving their transparency and reliability
- Our pre-audit is based on the methodology defined in version 3 of the IDI using the freely available data from the ITU website

Composite Indicators are powerful advocacy tools

COIN helps to ensure that composite indicators are … developed sensibly and used responsibly.

- ✓ '**Sensible development**' of a composite indicator implies a quality control process based on both conceptual and statistical considerations.
- ✓ '**Responsible use**' calls for care in drawing conclusions and recommendations without taking into account the conceptual context in which composite indicators were developed.



European Commission

European Commission's - Competence Centre on Composite Indicators and Scoreboards (JRC-COIN)

Support to EU Services in any policy area

Experience started in 2003

Centre Launched 02/2016

Support to international organisations

Methodology and guidelines

Handbook on Constructing

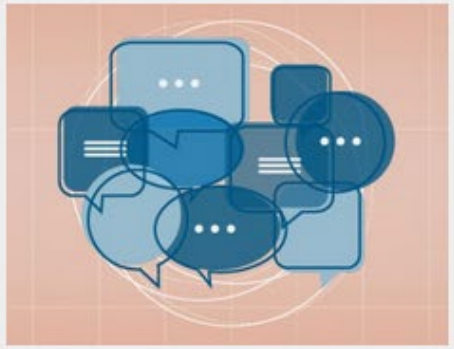
METHODOLOGY AND USER GUIDE

Training and Community of Practice

European Commission

Our Activity

Training



Community of Practice

Online platforms

Toolkit

Projects

Statistical Audits

COIN Open Days

CC-COIN Seminars and Workshops

https://knowledge4policy.ec.europa.eu/composite-indicators_en

Aim of the presentation

- Introduction to the methodology of Composite Indicators
- Assess how the IDI (v3) followed an accepted methodology
- With two main objectives:

1. Check the characteristics of the data and any potential limitations due to missing data and outliers
2. Check the statistical properties of aggregating indicators into the 2 pillars and into the overall ICT Development Index.

10 STEPS to build a Composite Indicator

Step 1 Conceptual framework

Step 2 Selection of indicators

Step 3 Data treatment

Step 4 Normalisation

Step 5 Weighting

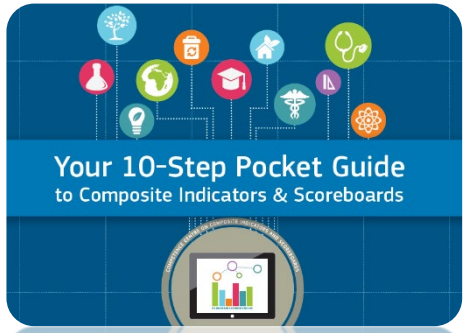
Step 6 Aggregation

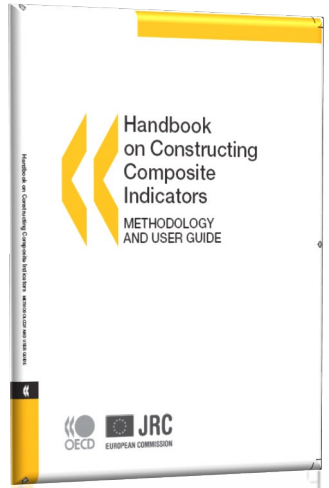
Step 7 Statistical coherence

Step 8 Robustness & Sensitivity

Step 9 Data sensemaking

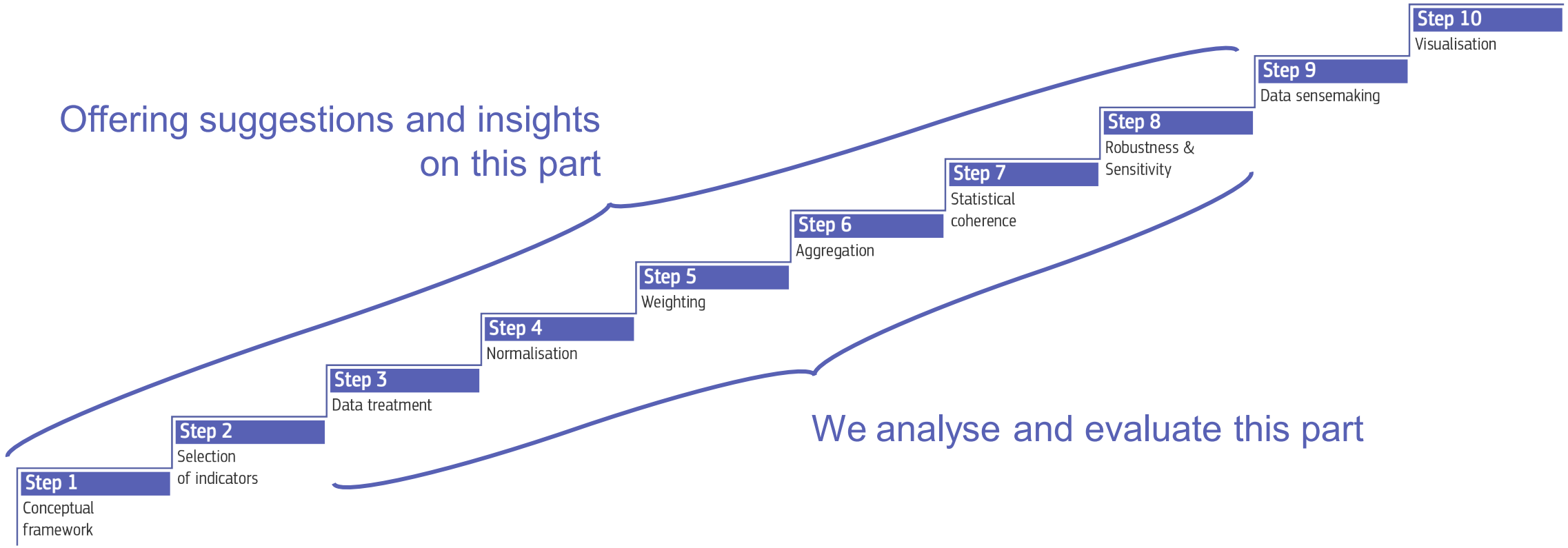
Step 10 Visualisation



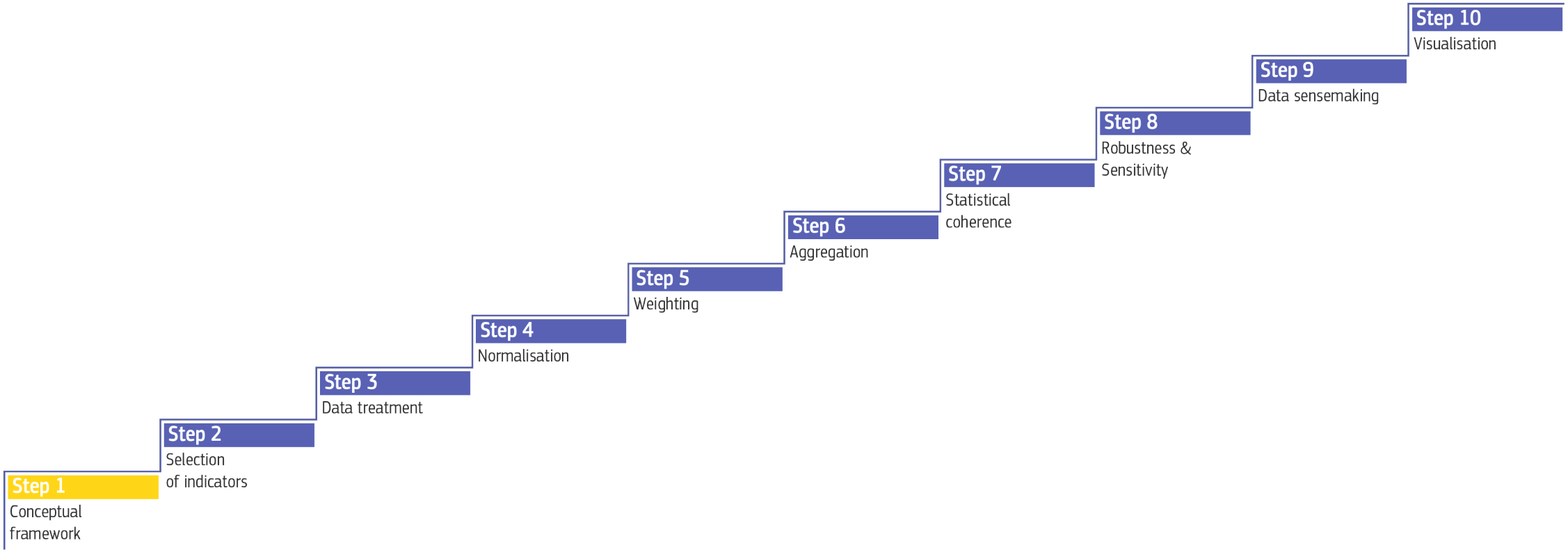


European Commission

Statistical assessment (by JRC-COIN)



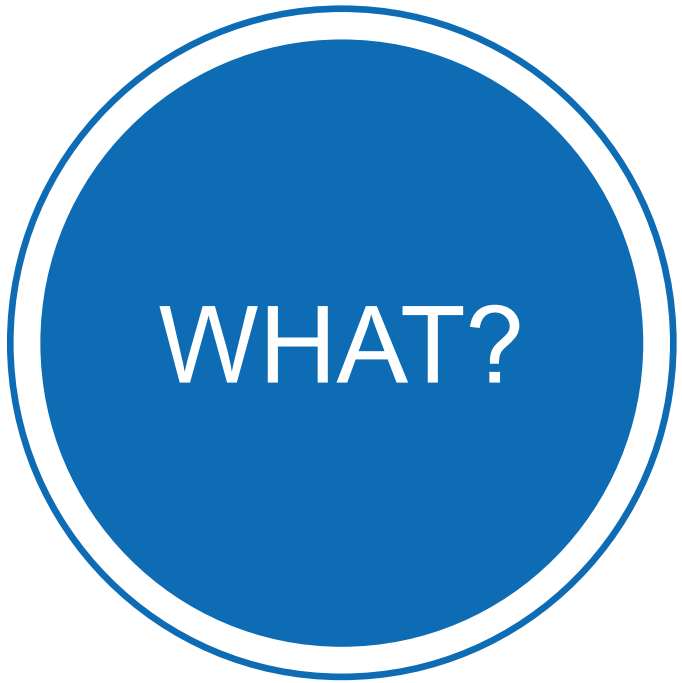
10 STEPS to build a Composite Indicator



Contribution by Laura di Bella

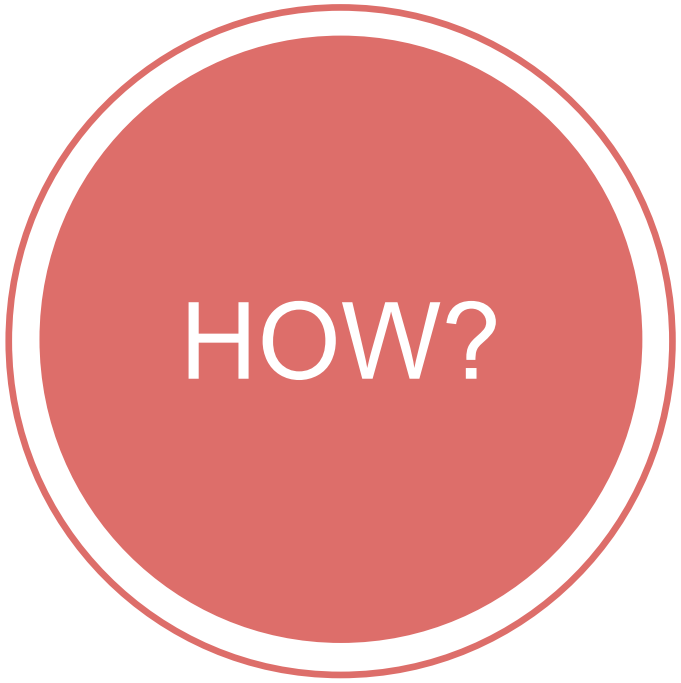
European Commission

Developing the conceptual framework



Measure Environmental Performance

Help National Governments refine their environmental policy agendas

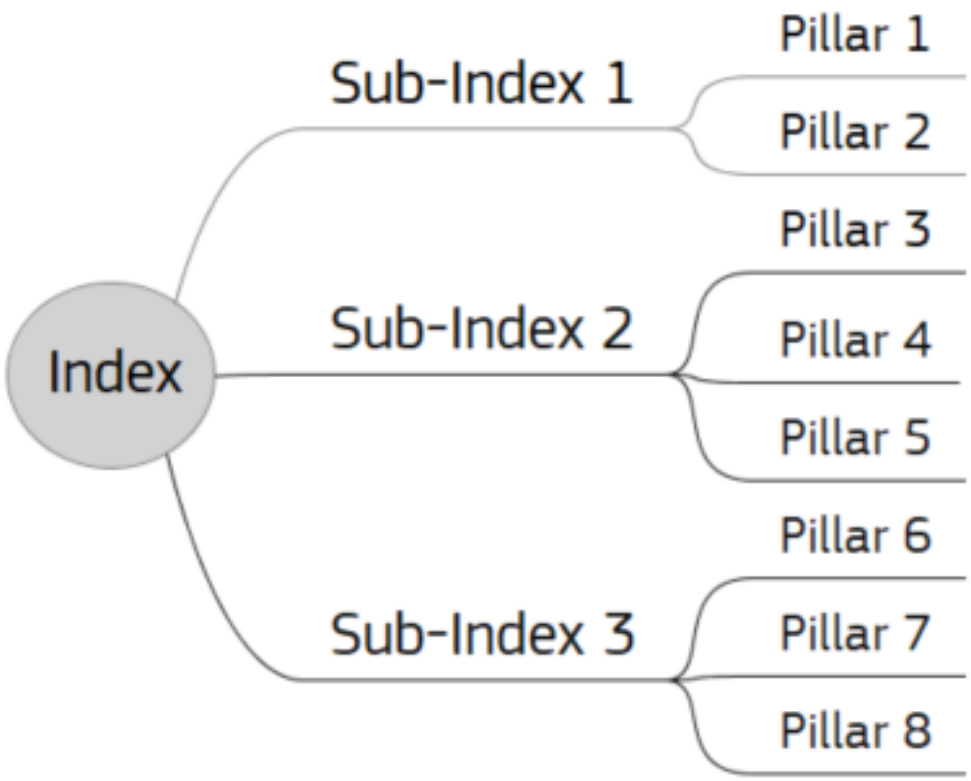


Decompose into issue categories (air, water, waste…) and identify indicators

Understand the domain

- Review the literature in the field
- Review existing initiatives, indicators, and indicator frameworks
- Engage with stakeholders – to better understand the what and the why
- Engage with experts – to better understand the what and the how

Structure the concept



The selected structure is an important part of the Index.

It affects the relative impact of the indicators on the Index.

The readability of the Index depends totally on it.

Step 1 in the IDI ✓

Table 1. Conceptual framework of the SDG Index.

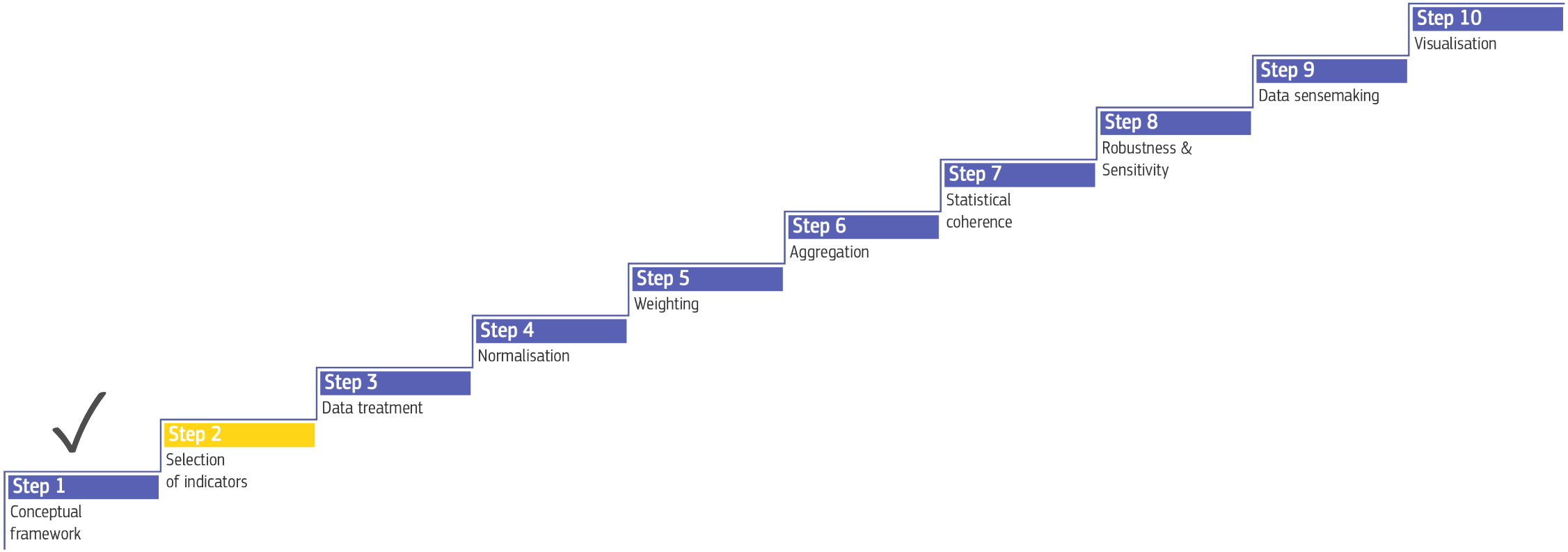
	Id	ITU Code	ICT Development Index (IDI)	Dir.
Universal Connectivity	1.1	i99H	Proportion of individuals who used the Internet	+
	1.2	xHH6_IDI	Proportion of households with Internet access at home	+
	1.3	i911mw	Active mobile-broadband subscriptions per 100 inhabitants	+
Meaningful Connectivity	2.1.1	i271G	Population covered by at least a 3G mobile network (%)[1]	+
	2.1.2	i271GA	Population covered by at least an 4G mobile network (%)[1]	+
	2.2	i136mwi_subs	Mobile broadband Internet traffic per mobile broadband subscriptions (GB)	+
	2.3	i135tfb_subs	Fixed broadband Internet traffic per fixed broadband subscriptions (GB)	+
	2.4	i271mb_high_ts_GNI	Mobile data and voice high-consumption basket price (as % of GNI per capita)	-
	2.5	i154_FBB_ts_GNI	Fixed-broadband Internet basket price (as % of GNI per capita)	-
	2.6	xHH18_IDI	Percentage of individuals owning a mobile phone	+

The structure is intuitive and clear. It is also well explained in the report.

The small number of indicators makes it easier to interpret.

The number of indicators is a measure of availability not of importance

10 STEPS to build a Composite Indicator



Contribution by Francesco Panella

European Commission

Selecting indicators

What makes a single indicator good?

Step 2 in the IDI ✓

Our recommendations

RELEVANT
SOUND
TIMELY
ACCESSIBLE
COMPARABLE

IDI's criteria

Relevant
Clear
Reliable
Useful as Measure
Timely
Available
High quality source

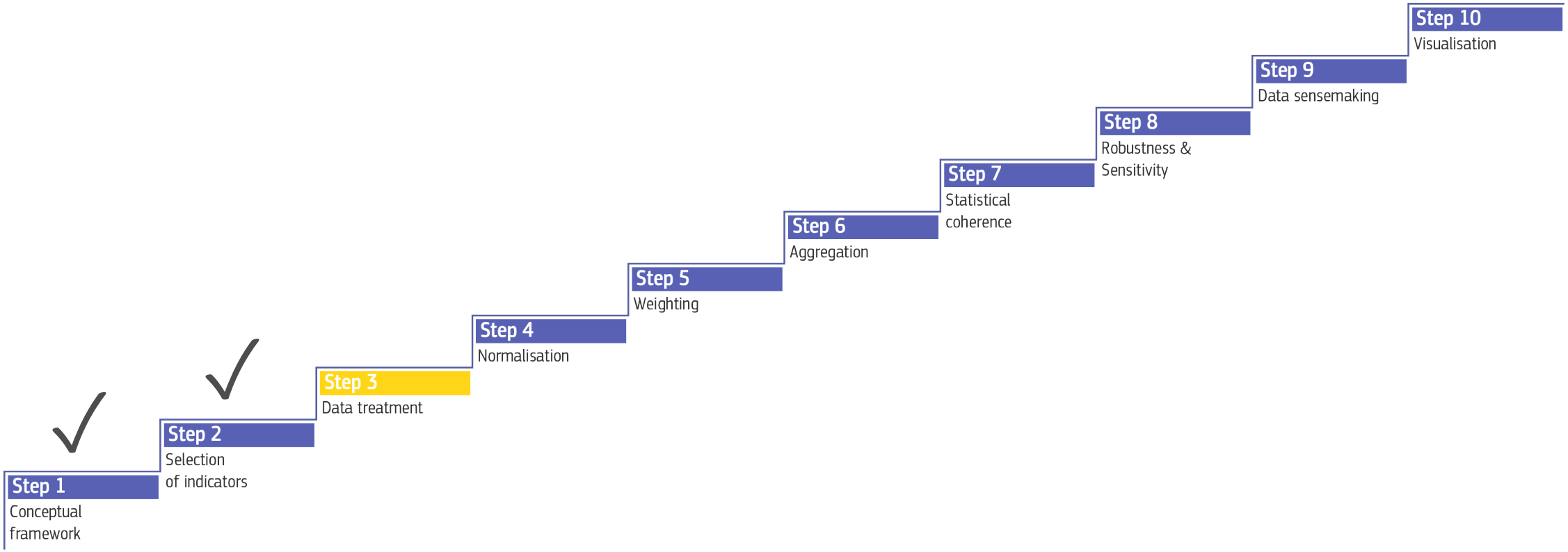
Additional suggestions

- Be careful to the denominators used
  (e.g. Population, GNI, GNI per capita)
  ✓ *all clear and intuitive in the IDI*

- Avoid the mix of Intensive and extensive measures
  (e.g. Number of Mobiles and Percentage of individuals with a mobile)
  ✓ *all intensive in the IDI*

- Avoid the mix of Stock data and Flow data
  (e.g. Population and pop. growth in the last year)
  ✓ *all stock in the IDI*

10 STEPS to build a Composite Indicator



Contribution by Oscar Smallenbroek

Data Treatment

Missing Data

Outliers

Definition of Missing data

- Missing values refer to the absence of data for one or more variables in a dataset.
- Missing values can occur for various reasons:
  - Non-response / Data entry errors / Deliberate omission / Not applicable

Practical tips

- Identify/reflect on the patterns of missing data
- Imputations often unreliable if data set contains more than 1/3 of missing values
  - **Indicator-level:** At least 80% of units should have valid data
  - **Unit-level:** At least 65-75% of the indicators for the unit within their aggregation levels should have valid data

Step 3 (Missing data) in the IDI ✓

Table 2. Summary statistics of the indicators included in the IDI.

ID	ITU Code	Number of observations	Missing data (%)	Mean	Skewness	Kurtosis	Min value	Max value	Lower Bound	Upper Bound
1.1	i99H	162	1.8	71.3	-1.0	0.0	5.8	100	0	95
1.2	xHH6_IDI	142	13.9	68.2	-0.9	-0.7	4.05	100	0	95
1.3	i911mw	165	0.0	84.8	1.1	3.4	2.61	285	0	146.8
2.1	i271	165	0.0	87	-1.8	2.6	17.5	100	0	100
2.2	i136mwi_subs*	164	0.6	3.97	-1.1	2.0	0	7.01	0	5.5
2.3	i135tfb_subs*	154	6.7	6.29	-0.8	-0.2	0	9.26	0	8.5
2.4	i271mb_high_ts_GNI	165	0.0	6.08	2.9	10.5	0.131	56.9	0	21.4
2.5	i154_FBB_ts_GNI	165	0.0	12.1	3.6	16.0	0.329	164	0	4.1
2.6	xHH18_IDI	158	4.2	82.6	-1.3	1.2	24.8	100	0	95

Indicators shaded in red have absolute skewness greater than 2.0 and kurtosis greater than 3.5.
** The values of these indicators are obtained after logarithmic transformation*
Source: European Commission's Joint Research Centre, 2023.

None of the indicators have a critical amount of missing values in the data tested.

Suggestion: Be careful to economies with less than 80% coverage within pillars
Comment: Using k-NN method is supported by JRC-COIN

Data treatment

Missing Data

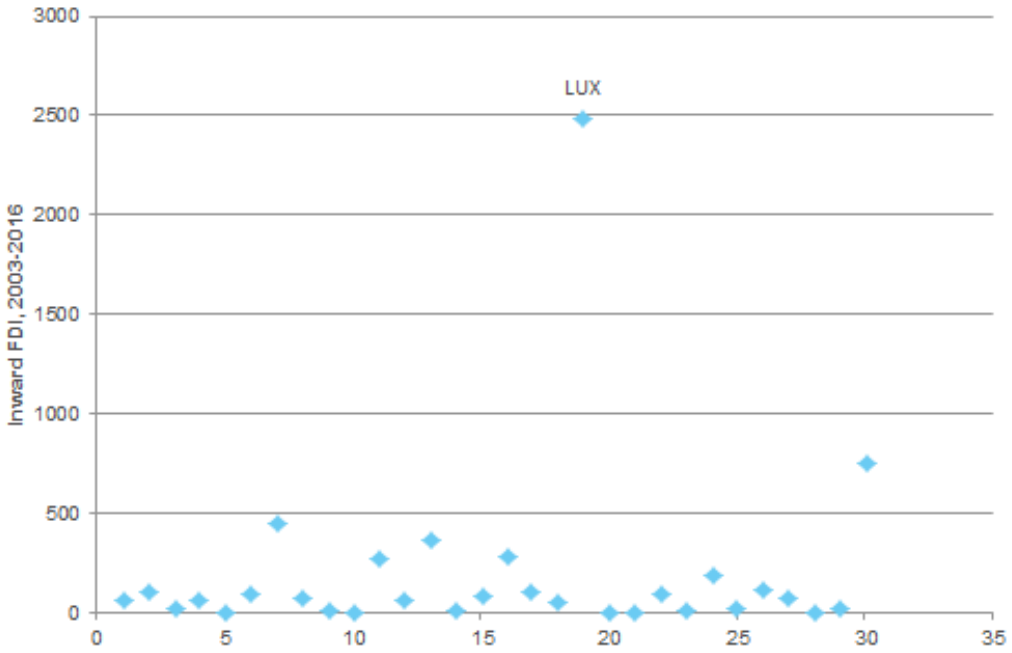
Outliers

Definition

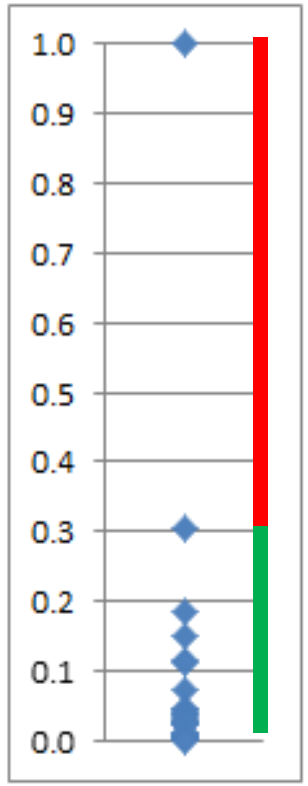
- ***Outlier-univariate***: an extreme value of an indicator, i.e. an observed value that deviates markedly or stands apart from the rest;

- Often they represent over(under)-performing units

What is the issue?



(min-max normalised data)



It may make the indicator irrelevant

Identification

- ***Easy rule of thumb:***

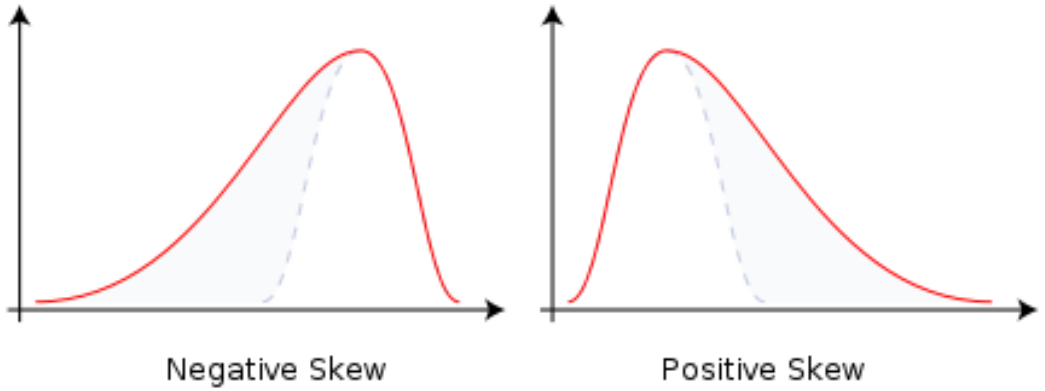
Check skewness and kurtosis (used by JRC-COIN)

- **Statistical Rules:**

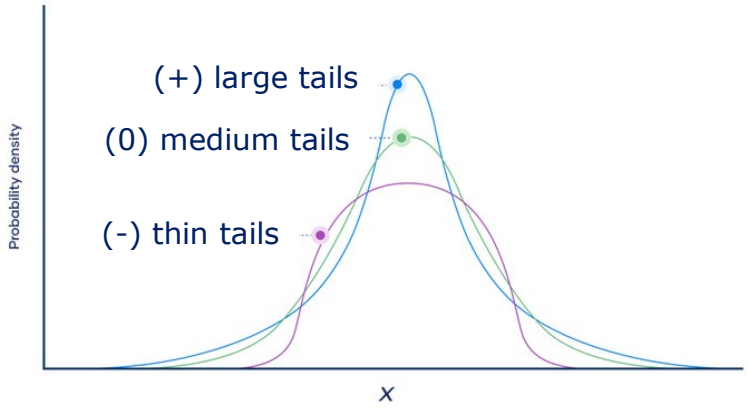
- Identify extreme z-scores
- Use interquartile range rule

Identification: skewness and kurtosis

Presence of outliers in a variable if **|skewness| > 2 & kurtosis > 3.5**



Skewness: measure of the asymmetry of a distribution



Kurtosis: measure of the weight of the tails relative to the centre of the distribution ("tailedness" of the distribution)

European Commission

Treatment/1 : winsorisation and capping

Winsorisation modifies outliers' values so to make them closer to other cases' values

- Values distorting the indicator distribution are ***replaced by the next value***.

- Winsorisation aims to mitigate the impact of extreme values by treating ***only potentially problematic observations***

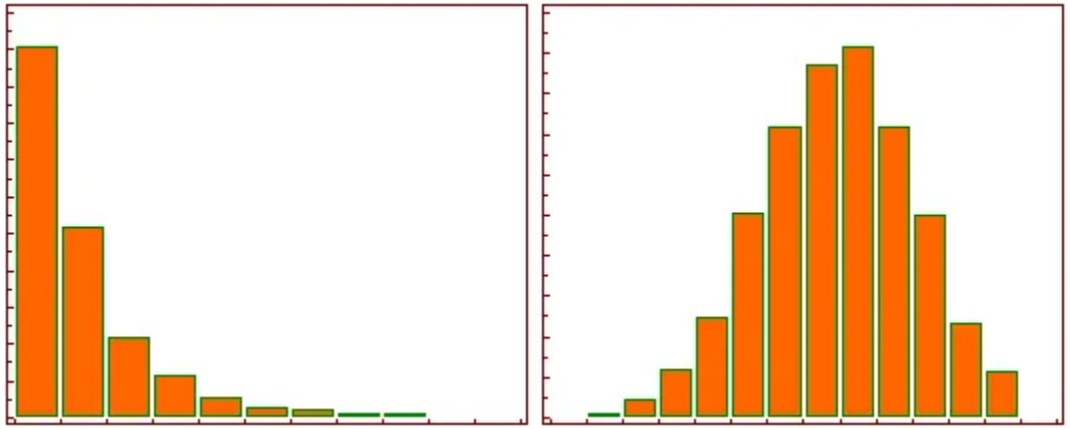
An informed winsorisation *could be achieved by defining thresholds of minimum and maximum for the indicators (See step 4 for the IDI)*

Treatment/2: transformations (logarithm)

Treat and transform **all the values** in the indicator

Recommended as an alternative to winsorisation **in case of identifying a high number of outliers (e.g. 5 or more)**

Particularly suited in case of Positive skewness (most common)



Raw Log-transformed

Step 3 (Outliers) in the IDI

Table 2. Summary statistics of the indicators included in the IDI.

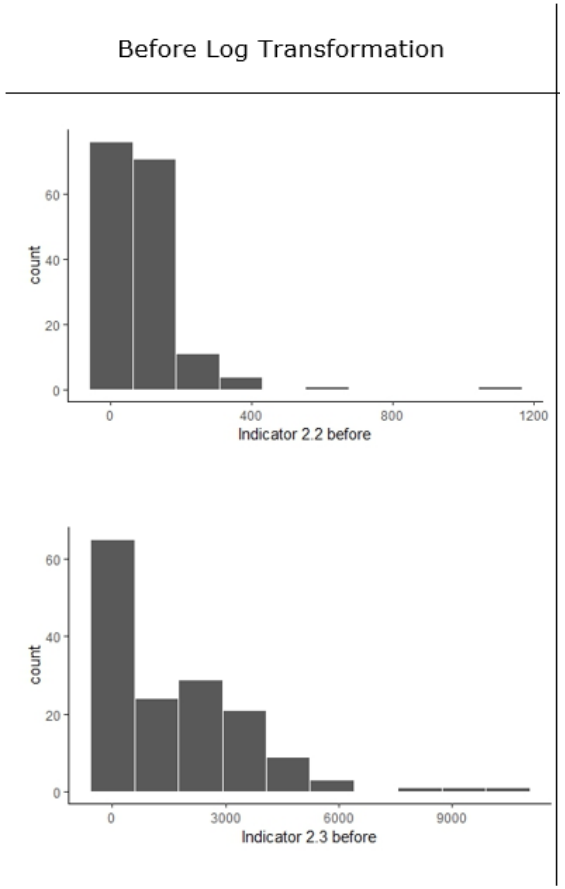
ID	ITU Code	Number of observations	Missing data (%)	Mean	Skewness	Kurtosis	Min value	Max value	Lower Bound	Upper Bound
1.1	i99H	162	1.8	71.3	-1.0	0.0	5.8	100	0	95
1.2	xHH6_IDI	142	13.9	68.2	-0.9	-0.7	4.05	100	0	95
1.3	i911mw	165	0.0	84.8	1.1	3.4	2.61	285	0	146.8
2.1	i271	165	0.0	87	-1.8	2.6	17.5	100	0	100
2.2	i136mwi_subs*	164	0.6	3.97	-1.1	2.0	0	7.01	0	5.5
2.3	i135tfb_subs*	154	6.7	6.29	-0.8	-0.2	0	9.26	0	8.5
2.4	i271mb_high_ts_GNI	165	0.0	6.08	2.9	10.5	0.131	56.9	0	21.4
2.5	i154_FBB_ts_GNI	165	0.0	12.1	3.6	16.0	0.329	164	0	4.1
2.6	xHH18_IDI	158	4.2	82.6	-1.3	1.2	24.8	100	0	95

Indicators shaded in red have absolute skewness greater than 2.0 and kurtosis greater than 3.5.
** The values of these indicators are obtained after logarithmic transformation*
Source: European Commission's Joint Research Centre, 2023.

*Indicators 2.2 and 2.3 were log-transformed, is it correct?

Step 3 (Outliers) in the IDI

Figure 1 - Distribution of indicators 2.2 and 2.3 before and after log-transformation



Source: Joint Research Centre of the European Commission analysis on ITU data, 2023

Step 3 (Outliers) in the IDI

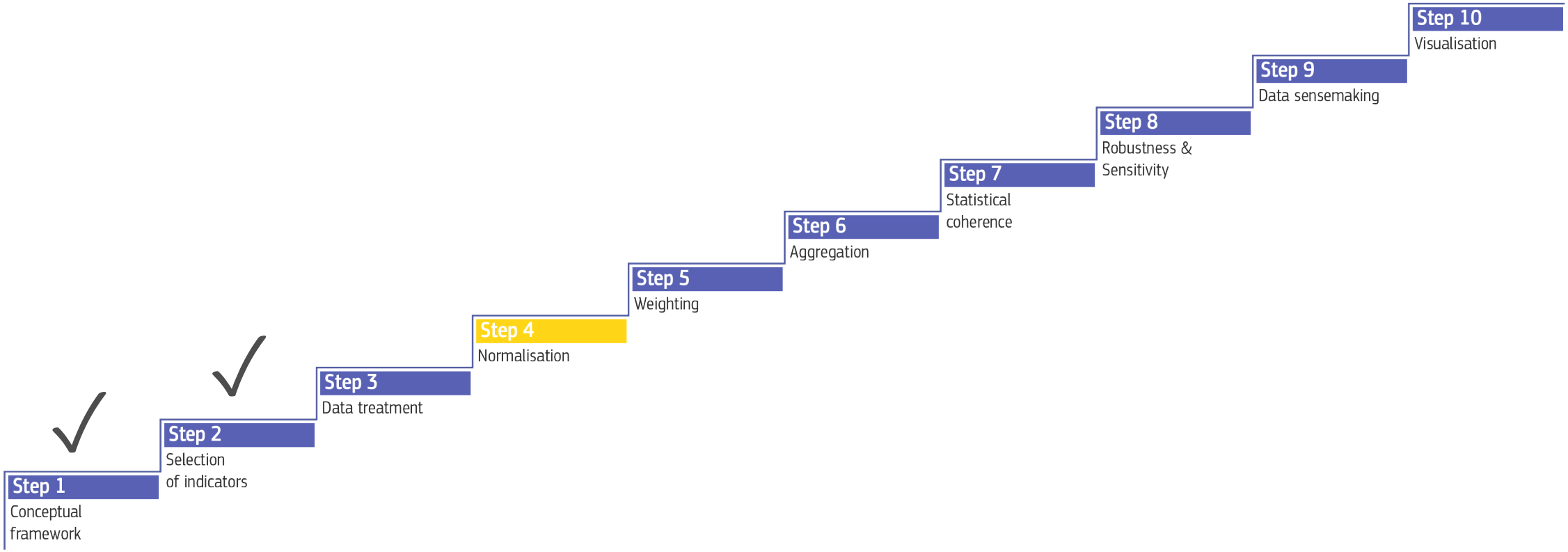
Table 2. Summary statistics of the indicators included in the IDI.

ID	ITU Code	Number of observations	Missing data (%)	Mean	Skewness	Kurtosis	Min value	Max value	Lower Bound	Upper Bound
1.1	i99H	162	1.8	71.3	-1.0	0.0	5.8	100	0	95
1.2	xHH6_IDI	142	13.9	68.2	-0.9	-0.7	4.05	100	0	95
1.3	i911mw	165	0.0	84.8	1.1	3.4	2.61	285	0	146.8
2.1	i271	165	0.0	87	-1.8	2.6	17.5	100	0	100
2.2	i136mwi_subs*	164	0.6	3.97	-1.1	2.0	0	7.01	0	5.5
2.3	i135tfb_subs*	154	6.7	6.29	-0.8	-0.2	0	9.26	0	8.5
2.4	i271mb_high_ts_GNI	165	0.0	6.08	2.9	10.5	0.131	56.9	0	21.4
2.5	i154_FBB_ts_GNI	165	0.0	12.1	3.6	16.0	0.329	164	0	4.1
2.6	xHH18_IDI	158	4.2	82.6	-1.3	1.2	24.8	100	0	95

Indicators shaded in red have absolute skewness greater than 2.0 and kurtosis greater than 3.5.
** The values of these indicators are obtained after logarithmic transformation*
Source: European Commission's Joint Research Centre, 2023.

Indicators 2.4 and 2.5 require attention, we will **get back to them** in Step 4.

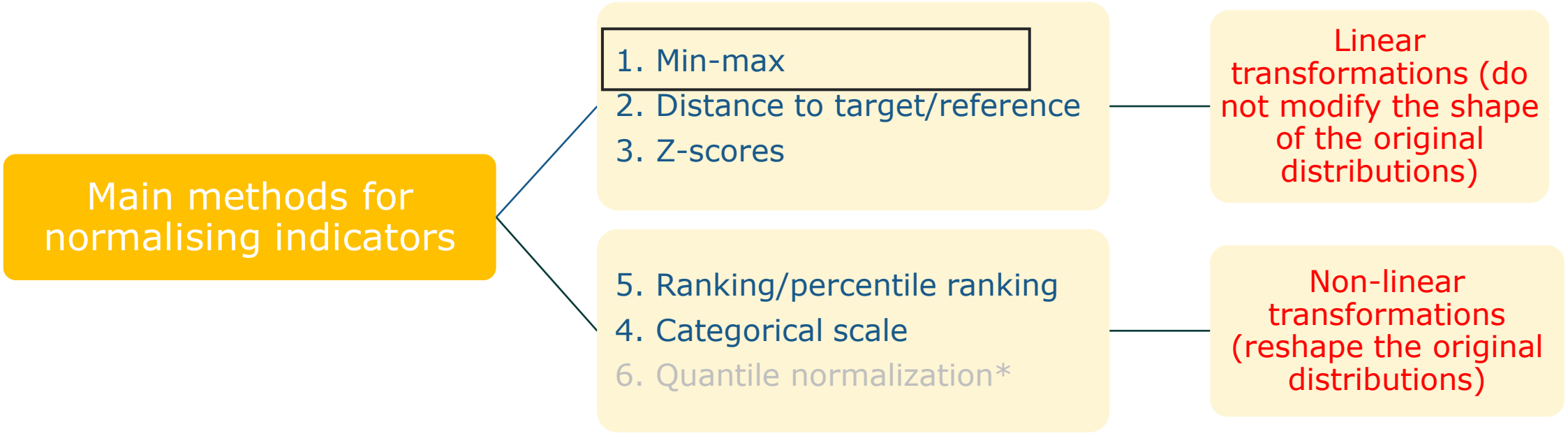
10 STEPS to build a Composite Indicator



Contribution by Marcos Dominguez-Torreiro

European Commission

Overview



Min-max (observed and defined)

MIN-MAX is the most common approach to normalise indicators.

It is based on the minimum and maximum **observed** in each indicator.

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

effects

- Unit of measurement
- μ, σ^2 not equal
- Variation range: [0, 1]
- Extreme values: no adjustments

✓ When the developers **defines meaningful references** and the population is very large and heterogeneous.

$$I = \frac{x - \text{low\_ref}(x)}{\text{high\_ref}(x) - \text{low\_ref}(x)}$$

Effects

- Unit of measurement
- μ, σ^2 not equal
- Variation range: [0, 1]
- Extreme values: treated

Step 4 in the IDI

Table 2. Summary statistics of the indicators included in the IDI.

ID	ITU Code	Number of observations	Missing data (%)	Mean	Skewness	Kurtosis	Min value	Max value	Lower Bound	Upper Bound
1.1	i99H	162	1.8	71.3	-1.0	0.0	5.8	100	0	95
1.2	xHH6_IDI	142	13.9	68.2	-0.9	-0.7	4.05	100	0	95
1.3	i911mw	165	0.0	84.8	1.1	3.4	2.61	285	0	146.8
2.1	i271	165	0.0	87	-1.8	2.6	17.5	100	0	100
2.2	i136mwi_subs*	164	0.6	3.97	-1.1	2.0	0	7.01	0	5.5
2.3	i135tfb_subs*	154	6.7	6.29	-0.8	-0.2	0	9.26	0	8.5
2.4	i271mb_high_ts_GNI	165	0.0	6.08	2.9	10.5	0.131	56.9	0	21.4
2.5	i154_FBB_ts_GNI	165	0.0	12.1	3.6	16.0	0.329	164	0	4.1
2.6	xHH18_IDI	158	4.2	82.6	-1.3	1.2	24.8	100	0	95

Indicators shaded in red have absolute skewness greater than 2.0 and kurtosis greater than 3.5.
** The values of these indicators are obtained after logarithmic transformation*
Source: European Commission's Joint Research Centre, 2023.

MIN-MAX method with defined thresholds is applied in the IDI

Step 4 in the IDI ✓

Table 3. Summary statistics of the indicators after normalisation

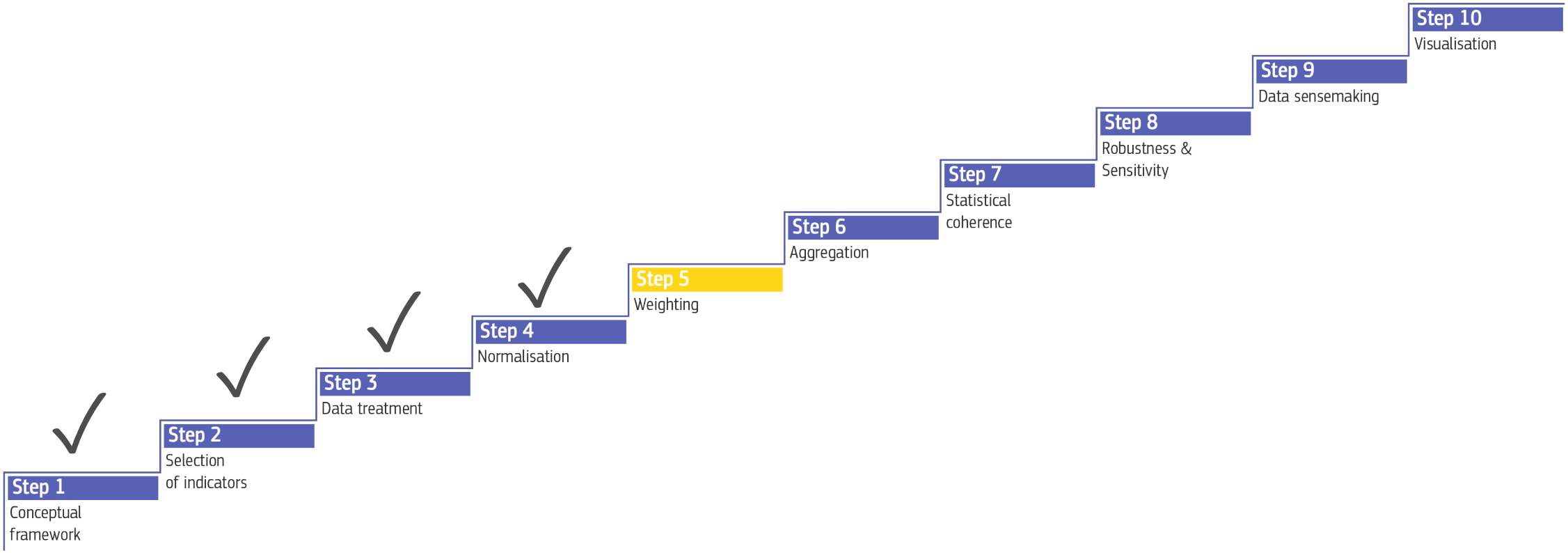
ID	ITU Code	Mean	Skewness	Kurtosis	Minimum value	Maximum value
1.1	i99H	73.0	-1.0	0.0	0	100
1.2	xHH6_IDI	70.1	-0.9	-0.7	0	100
1.3	i911mw	55.1	-0.1	-0.6	0	100
2.1	i271	84.3	-1.8	2.6	0	100
2.2	i136mwi_subs*	63.8	-1.1	2.0	0	100
2.3	i135tfb_subs*	68.3	-0.8	-0.2	0	100
2.4	i271mb_high_ts_GNI	78.1	-1.5	1.2	0	100
2.5	i154_FBB_ts_GNI	83.4	-2.1	3.5	0	100
2.6	xHH18_IDI	81.2	-1.4	1.5	0	100

Indicators shaded in red have absolute skewness greater than 2.0 and kurtosis greater than 3.5.
** The values of these indicators are obtained after logarithmic transformation*
Source: European Commission's Joint Research Centre, 2023.

Indicators 2.4 and 2.5 issues related to outliers are solved by the normalisation approach.

JRC-COIN do not suggest further transformations for 2.5.

10 STEPS to build a Composite Indicator



Contribution by Matija Kovacic

Weighting in composite indicators

- **Meaning of weighting**:
  - Trade-off between indicators/pillars/sub-pillars ("*implicit importance*")

- **Selecting a weighting scheme** is not a simple task:
  - There is no "*one-size-fits-all*" solution
  - Stakeholders may have different opinions on choosing weighting scheme

- **A composite is your own product**: theoretical framework + transparency

Approaches to setting weights

- **Equal weights** (vast majority)
- **Weighting based on statistical methods**
- **Weights based on public/expert opinion**

Approaches to setting weights: examples

Composite Indicator	Weighting scheme
Human Development Index	**Equal weights**
Multidimensional Poverty Index	**Equal weights** **Expert opinion**
Quality of Life Index	**Equal weights**
Better Life Index	**Equal weights**
Social Progress Index	**Principal component analysis**
Corruption Perception Index	**Equal weights**
Rule of Law Index	**Equal weights** **Expert opinion**
Environmental Performance Index	**Expert opinion**
Gender Equality Index	**Expert opinion**
ICT Development Index 2023 (v3)	**Equal Weights**

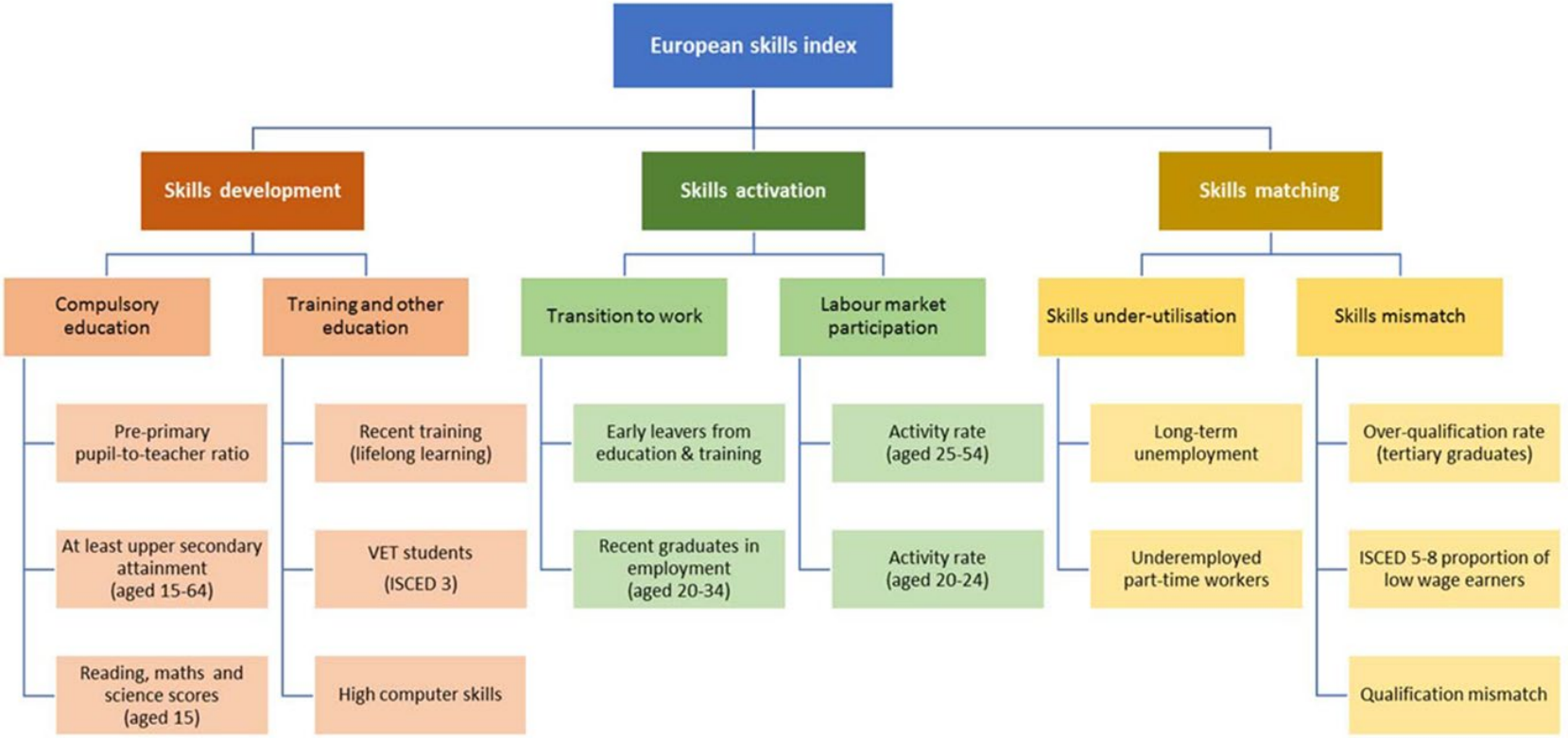
Equal weighting

- Equal weighting is *the most common* scheme appearing in composite indicators
- **Justifications** of choosing equal weights:
  - *Simple and easy to communicate*
  - *No agreement between stakeholders*

However, equal weighting …

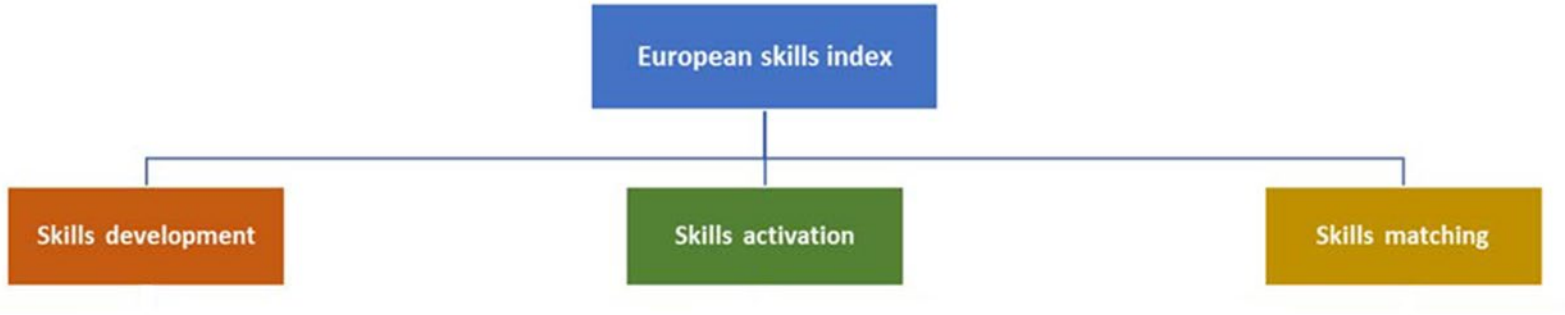
- **<u>does not mean</u> not distributing weights at all**
- **<u>does not mean</u> equal "contribution" of the indicators to the composite**

Example: European Skills Index



Source: *European Skills Index (2020), Cedefop.*

Example: European Skills Index



Equal Weights

Pillars	Pearson Correlation Coefficient	R^2
Skills Development	0.80	0.64
Skills Activation	0.81	0.66
Skills Matching	0.64	0.41

Step 5 in the IDI ✓

Table 6: *Correlations between the pillars and with the index*

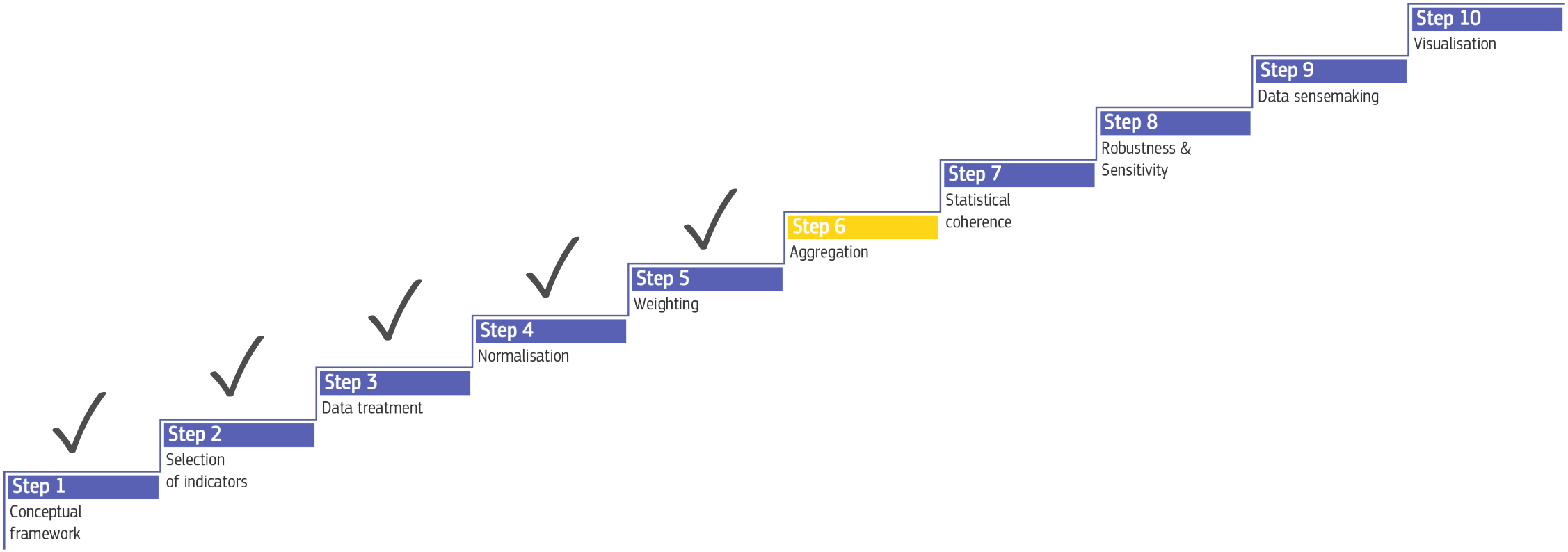
	Universal Connectivity	Meaningful Connectivity	IDI
Universal Connectivity		0.88	**0.98**
Meaningful Connectivity	0.88		**0.96**
IDI	**0.98**	**0.96**	

Note: Numbers represent the Pearson correlation coefficients. Good correlations (i.e. Pearson correlation coefficients greater than 0.30 and lower than 0.92) are highlighted in green. Correlations with low values (here <0.30) are in grey. Correlations at risk of redundancy (here >0.92) are highlighted in blue.

Source: European Commission's Joint Research Centre, 2023.

Differently from the example, the pillars of the IDI are correlated with the IDI in a balanced way. From this perspective, there is no need to weight them differently

10 STEPS to build a Composite Indicator



European Commission

Arithmetic mean

The arithmetic mean of a list of n real numbers equals:

$$\frac{1}{n}\sum_{i=1}^{n} x_i \quad = \quad \frac{x_1 + x_2 + \cdots + x_n}{n}$$

This is the simplest, most intuitive and most widespread aggregation method

Perfect (and constant) substitutability – underperformance in one component can be compensated by equivalent overperformance in another

Geometric mean

The ***geometric mean*** of a list of *n positive* real numbers equals:

$$\sqrt[n]{\prod_{i=1}^{n} x_i} = \sqrt[n]{x_1 \times x_2 \times \cdots \times x_n}$$

The first “less-compensatory” option

Partial substitutability – *unbalanced performance is always penalised* by the aggregation formula when compared to arithmetic aggregation

Step 6 in the IDI ✓

Table 6: *Correlations between the pillars and with the index*

	Universal Connectivity	Meaningful Connectivity	IDI
Universal Connectivity		0.88	**0.98**
Meaningful Connectivity	0.88		**0.96**
IDI	**0.98**	**0.96**	

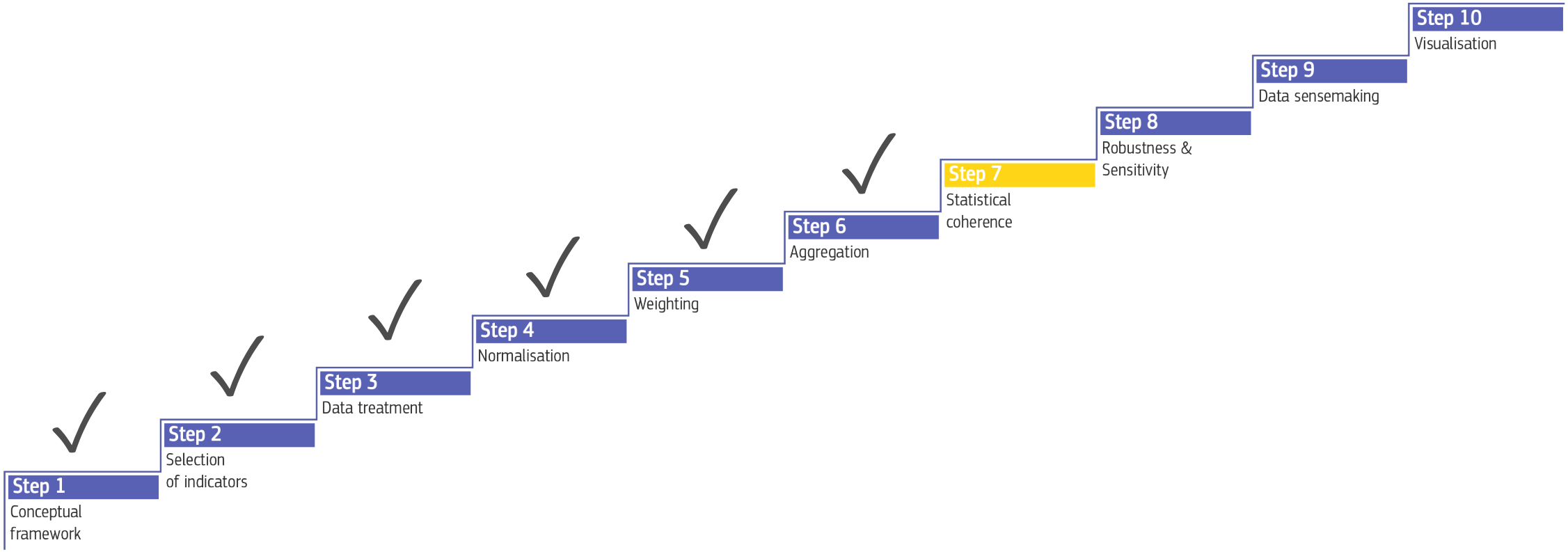
Note: Numbers represent the Pearson correlation coefficients. Good correlations (i.e. Pearson correlation coefficients greater than 0.30 and lower than 0.92) are highlighted in green. Correlations with low values (here <0.30) are in grey. Correlations at risk of redundancy (here >0.92) are highlighted in blue.

Source: European Commission's Joint Research Centre, 2023.

The effect of the Geometric Average is stronger in composite indicators with low correlations.

In the IDI the two methods would not differ dramatically.

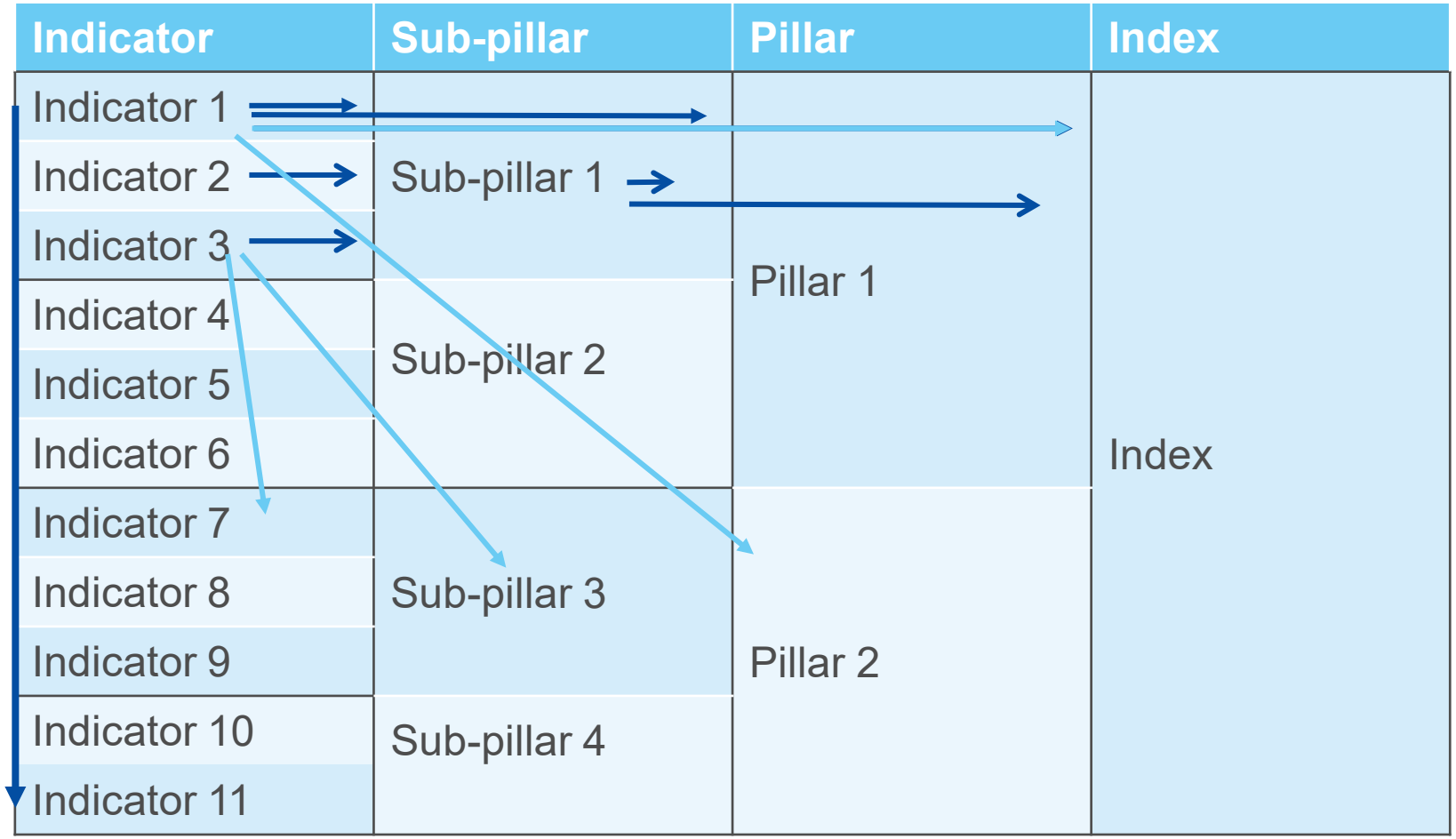
10 STEPS to build a Composite Indicator



Contribution by Eleni Papadimitriou

European Commission

Which correlations do we observe?



Check whether indicators:

- Are negatively related corr < -0.3
- Are under-represented
  -0.3 < x < 0.3
- Dominate the framework

What responses we can find?

- Are indicators allocated in the **right dimension**?
- Are some indicators **over or under represented** in the aggregate **Index**?
- Up to what **level should we aggregate**?

Step 7 in the IDI – Between indicators ✓

Table 4: Correlations between indicators

ITU Code	ID	1.1	1.2	1.3	2.1	2.2	2.3	2.4	2.5	2.6
i99H	1.1		0.90	0.72	0.69	0.59	0.34	0.78	0.78	0.90
xHH6_IDI	1.2	0.90		0.74	0.69	0.60	0.45	0.75	0.74	0.81
i911mw	1.3	0.72	0.74		0.64	0.51	0.30	0.67	0.61	0.70
i271	2.1	0.69	0.69	0.64		0.45	0.36	0.70	0.64	0.65
i136mwi_subs	2.2	0.59	0.60	0.51	0.45		0.32	0.56	0.56	0.52
i135tfb_subs	2.3	0.34	0.45	0.30	0.36	0.32		0.34	0.24	0.27
i271mb_high_ts_GNI	2.4	0.78	0.75	0.67	0.70	0.56	0.34		0.78	0.72
i154_FBB_ts_GNI	2.5	0.78	0.74	0.61	0.64	0.56	0.24	0.78		0.76
xHH18_IDI	2.6	0.90	0.81	0.70	0.65	0.52	0.27	0.72	0.76	

Note: Numbers represent the Pearson correlation coefficients. Good correlations (i.e. Pearson correlation coefficients greater than 0.30 and lower than 0.92) are highlighted in green. Correlations with low values (here <0.30) are written in grey. Correlations at risk of redundancy (here >0.92) are highlighted in blue.

Source: European Commission's Joint Research Centre, 2023.

The correlation structure is strong and robust

Only exception Indicator 2.3, but far from critical

Step 7 in the IDI – Indicators with Aggregates ✓

Table 5. *Correlations between indicators and their aggregates*

ITU Code	ID	Universal Connectivity	Meaningful Connectivity	IDI
i99H	1.1	**0.94**	0.87	**0.94**
xHH6_IDI	1.2	**0.96**	0.84	**0.93**
i911mw	1.3	0.88	0.73	0.84
i271	2.1	0.72	0.80	0.78
i136mwi_subs	2.2	0.61	0.72	0.68
i135tfb_subs	2.3	0.41	0.54	0.48
i271mb_high_ts_GNI	2.4	0.79	0.90	0.86
i154_FBB_ts_GNI	2.5	0.76	0.87	0.83
xHH18_IDI	2.6	0.87	0.84	0.88

Note: Numbers represent the Pearson correlation coefficients. Good correlations (i.e. Pearson correlation coefficients greater than 0.30 and lower than 0.92) are highlighted in green. Correlations with low values (here <0.30) are in grey. Correlations at risk of redundancy (here >0.92) are highlighted in blue.

Source: European Commission's Joint Research Centre, 2023.

Indicators correlating very well with pillars (better with own pillars)

First two indicators are strongly correlated with the aggregates, but not to the point of damaging Ind 1.3 representation.

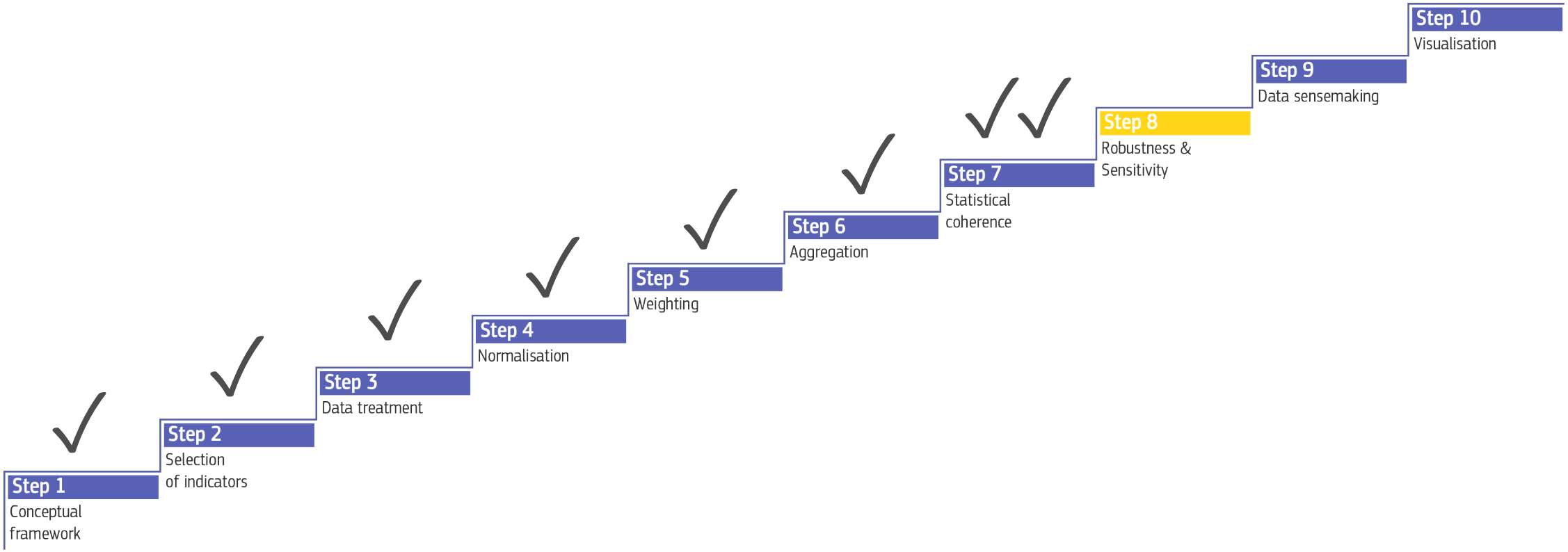
Step 7 in the IDI - Between aggregates ✓

Table 6: *Correlations between the pillars and with the index*

	Universal Connectivity	Meaningful Connectivity	IDI
Universal Connectivity		0.88	**0.98**
Meaningful Connectivity	0.88		**0.96**
IDI	**0.98**	**0.96**	

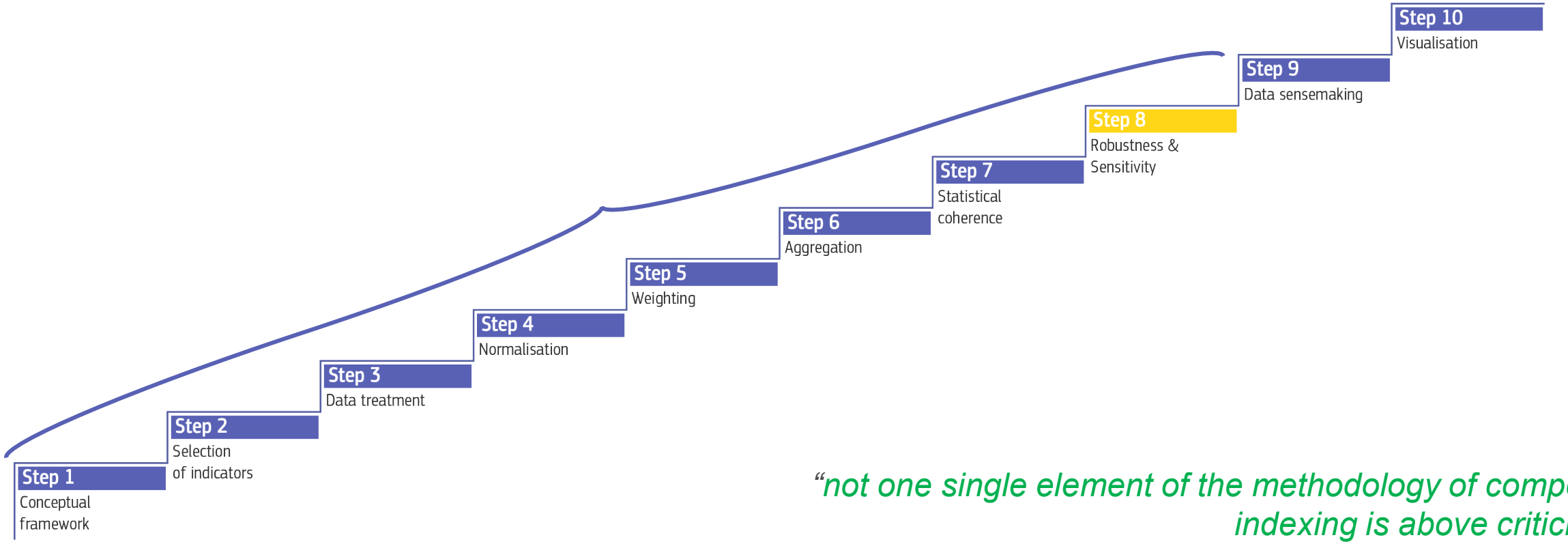
The pillars are well correlated and not conflicting, they are also represented in a balanced way.

10 STEPS to build a Composite Indicator



Contribution by William Becker

Which steps in the construction of a composite indicator are uncertain?



"not one single element of the methodology of composite indexing is above criticism".
Booysen (2002, p.131)

European Commission

Step 8 in the IDI

The JRC-COIN studied the effect of two choices:

1. Log-transformation of indicators 2.2. and 2.3 (Mobile and Fixed traffic)
2. Aggregation in two pillars vs direct aggregation

Step 8 in the IDI

1. Log-transformation of indicators 2.2. and 2.3 (Mobile and Fixed traffic)
2. Aggregation in two pillars vs direct aggregation

Table 5. *Correlations between indicators and their aggregates*

ITU Code	ID	Universal Connectivity	Meaningful Connectivity	IDI
i99H	1.1	**0.94**	0.87	**0.94**
xHH6_IDI	1.2	**0.96**	0.84	**0.93**
i911mw	1.3	0.88	0.73	0.84
i271	2.1	0.72	0.80	0.78
i136mwi_subs	2.2	0.61	0.72	0.68
i135tfb_subs	2.3	0.41	0.54	0.48
i271mb_high_ts_GNI	2.4	0.79	0.90	0.86
i154_FBB_ts_GNI	2.5	0.76	0.87	0.83
xHH18_IDI	2.6	0.87	0.84	0.88

Table 8: *Correlations between indicators, pillars and IDI (2.2 and 2.3 not transformed)*

ITU Code	ID	Universal Connectivity	Meaningful Connectivity	IDI
i99H	1.1	**0.94**	0.86	**0.94**
xHH6_IDI	1.2	**0.96**	0.83	**0.93**
i911mw	1.3	0.88	0.74	0.85
i271	2.1	0.72	0.78	0.77
i136mwi_subs	2.2	0.55	0.68	0.63
i135tfb_subs	2.3	0.37	0.55	0.46
i271mb_high_ts_GNI	2.4	0.79	0.87	0.85
i154_FBB_ts_GNI	2.5	0.76	0.84	0.82
xHH18_IDI	2.6	0.87	0.81	0.87

Step 8 in the IDI

1. Log-transformation of indicators 2.2. and 2.3 (Mobile and Fixed traffic)
2. Aggregation in two pillars vs direct aggregation

Table 10: *Correlations between indicators and IDI (no pillars)*

ITU Code	ID	IDI
i99H	1.1	**0.93**
xHH6_IDI	1.2	**0.91**
i911mw	1.3	0.82
i271	2.1	0.80
i136mwi_subs	2.2	0.70
i135tfb_subs	2.3	0.50
i271mb_high_ts_GNI	2.4	0.88
i154_FBB_ts_GNI	2.5	0.85
xHH18_IDI	2.6	0.88

The correlations change is very low (<0.03).

According to correlation analysis this second option is also an acceptable choice

Important

Having no intermediate aggregation would give to the second group twice the weight (six indicators vs three), reducing the balance we saw before.

Pre-Audit of IDI v3 - Conclusions 1

- The chosen structure, included the two levels, is clear and intuitive, and is well justified in the methodological report, it improves the transparency
- The small number of indicators and the good selection criteria, enhance the readability of the index and makes it easy to scrutinise
- Data coverage is very good, and outliers are well managed with an informed approach
- The choice of equal weights and arithmetic mean fit well with the structure
- The correlation of only one indicator (Fixed broadband Internet traffic per fixed broadband subscriptions (GB)) is a bit weak in some cases, but not at a relevant level
- The correlation structure is strong and positive, up to act as a strength of the IDI

Pre-Audit of IDI v3 - Conclusions 2

According to our analysis the IDI is a reliable tool, and the framework has a good statistical coherence that ensures a robust structure.

The pre-audit also acknowledges the significant efforts of the developers to obtain a balanced and transparent result.

Thank you

Our email: JRC-COIN@ec.europa.eu

Link to our activities and the training Here

European Commission

Imputation of Missing data
k-Nearest Neighbors (kNN)

Imputation method, based on using the information of available indicators to find the most similar peers and use their value to impute the missing

PROS

- Intuitive compared to other multivariate approaches.
- Imputes possible values
- Easy to implement with statistical software

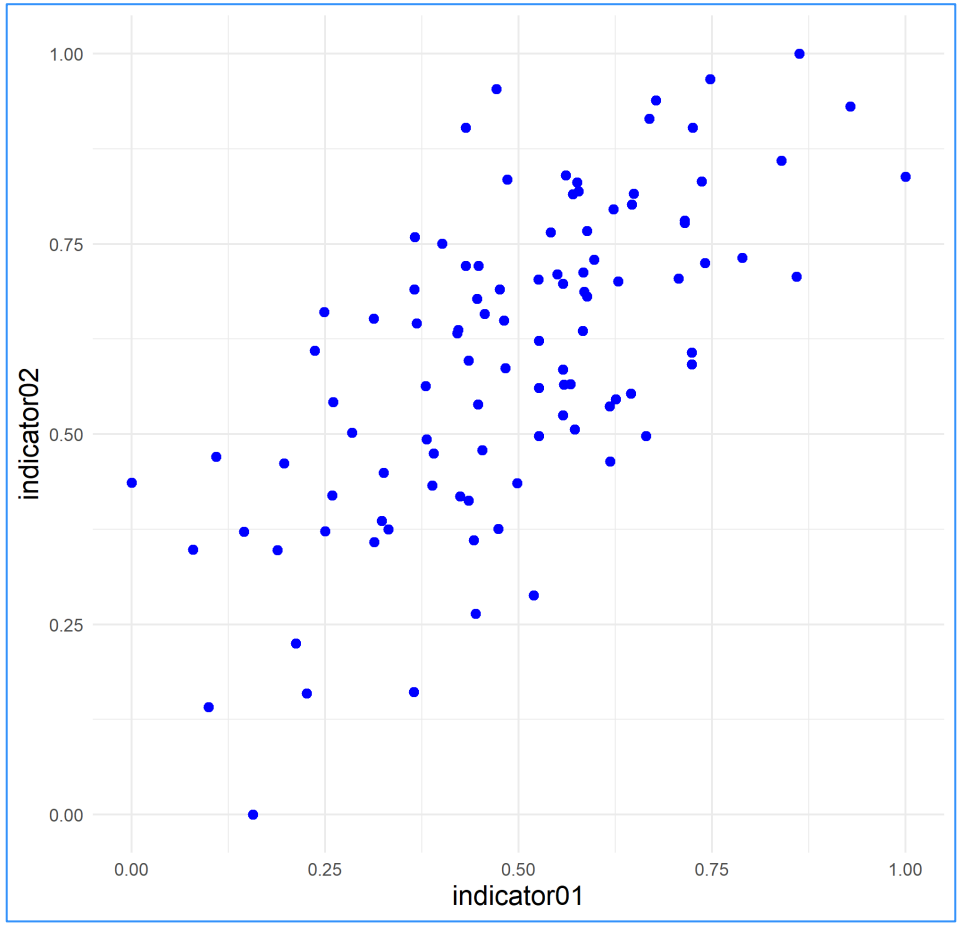
CONS

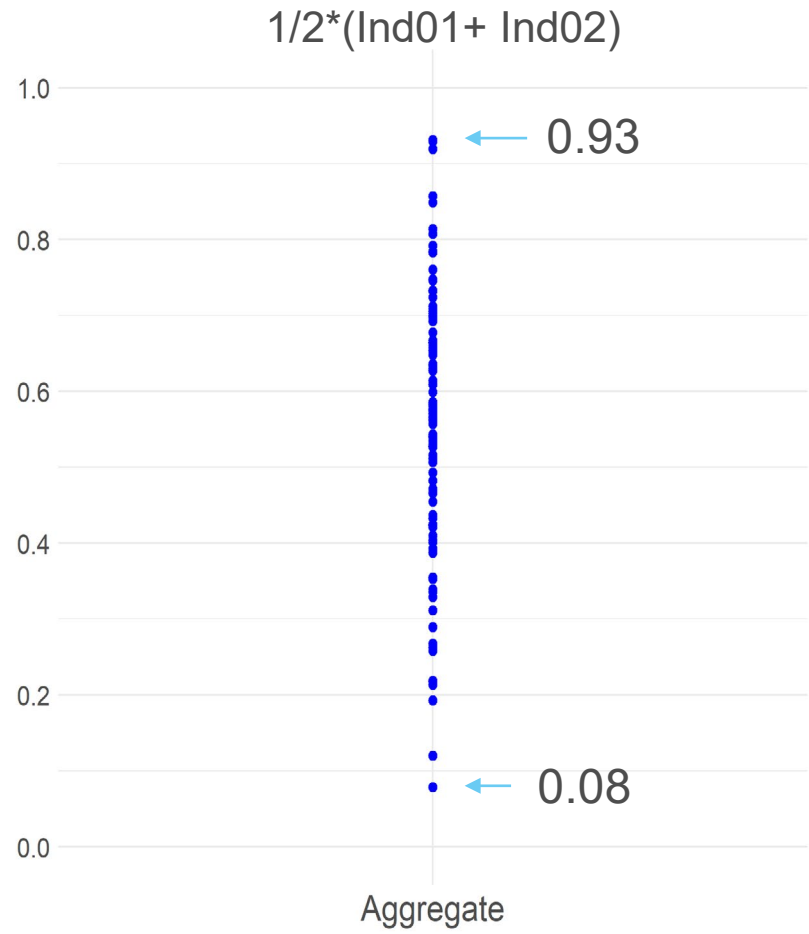
- Needs a large
- Needs a good set of characteristics to compare cases
- Needs statistical software

Why correlations are important?

Indicator 1
Indicator 2

Correlation
0.663



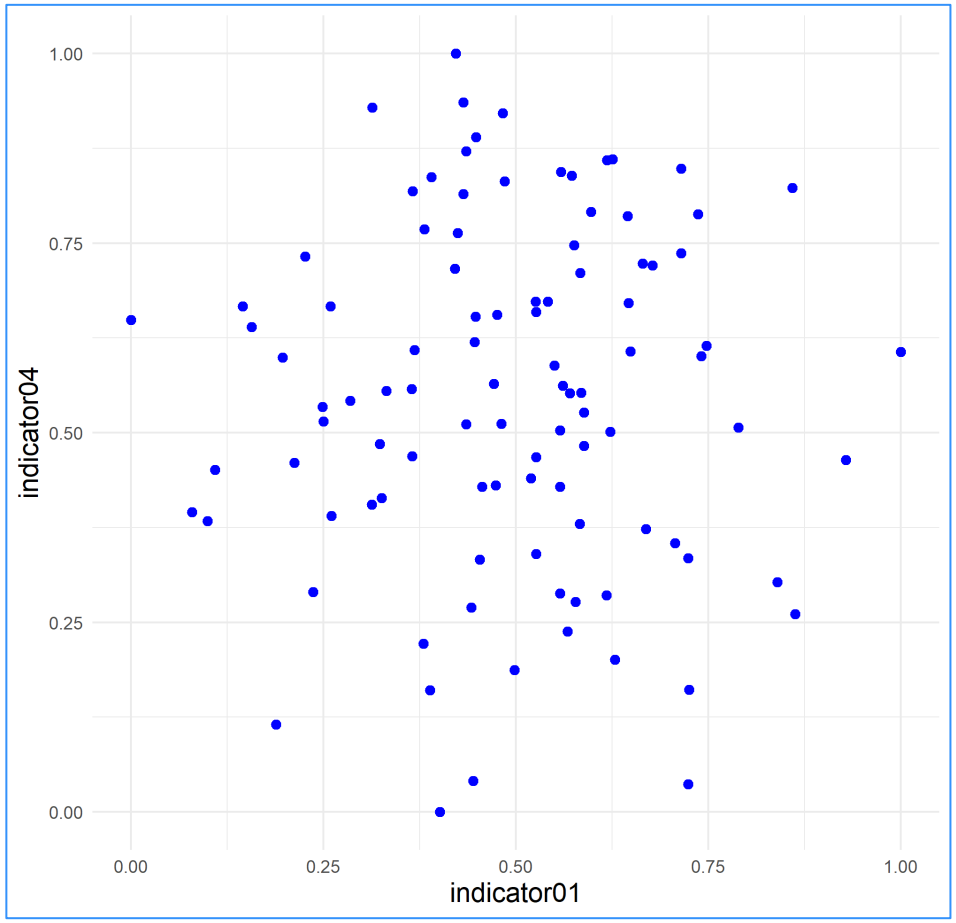


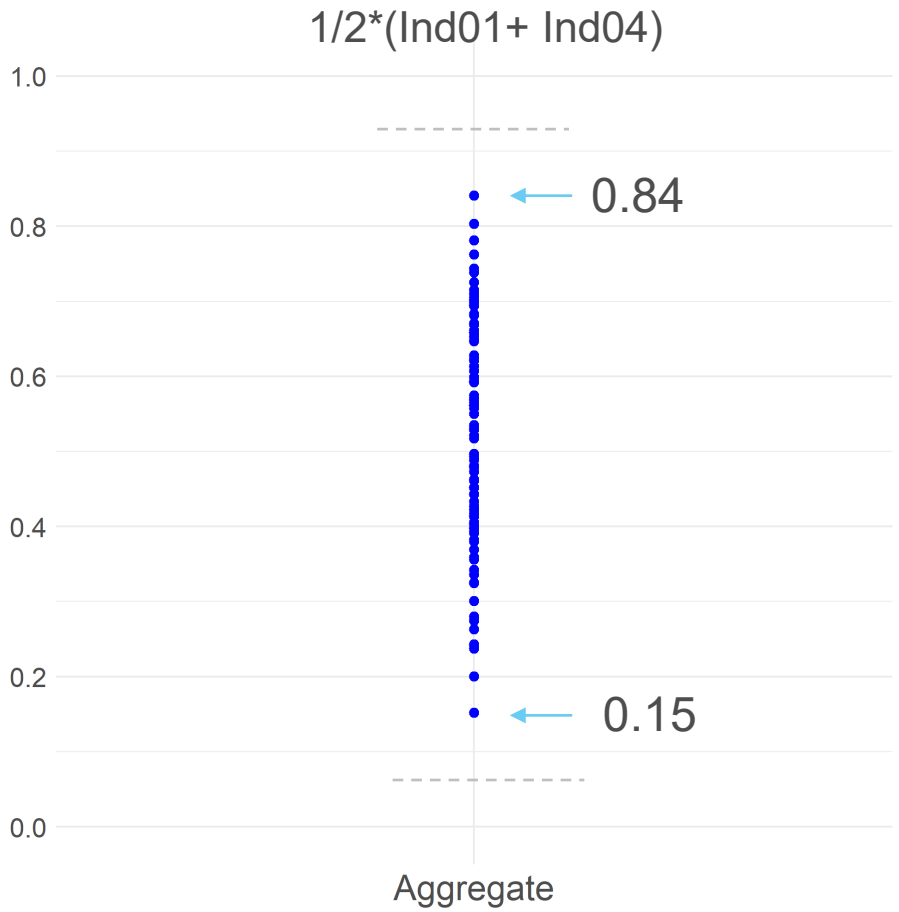
European Commission

Why correlations are important?

Indicator 1
Indicator 4

Correlation
0.011



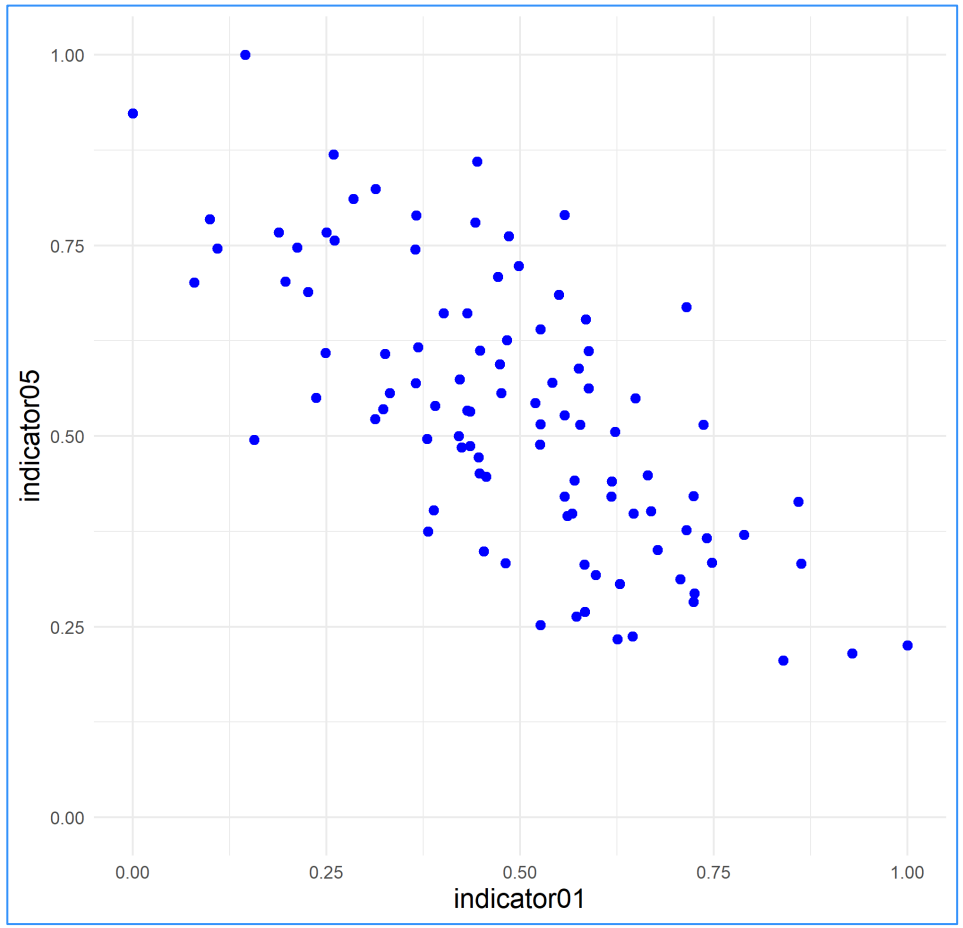


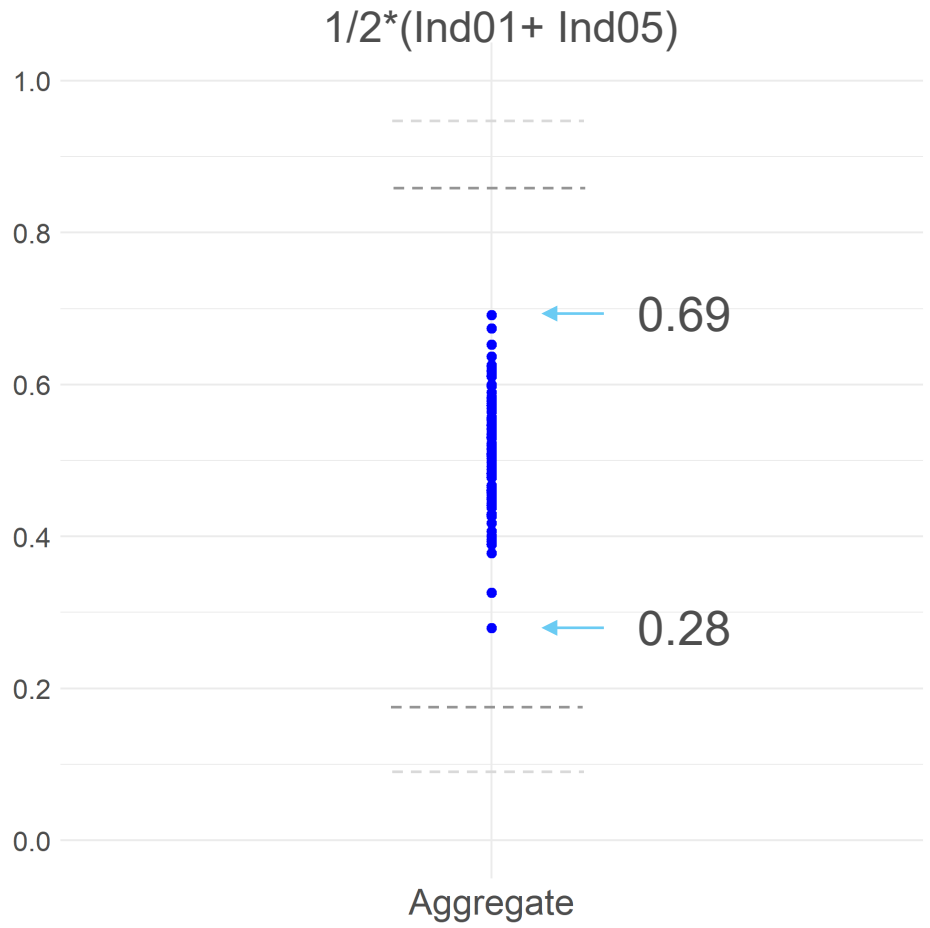
European Commission

Why correlations are important?

Indicator 1
Indicator 5

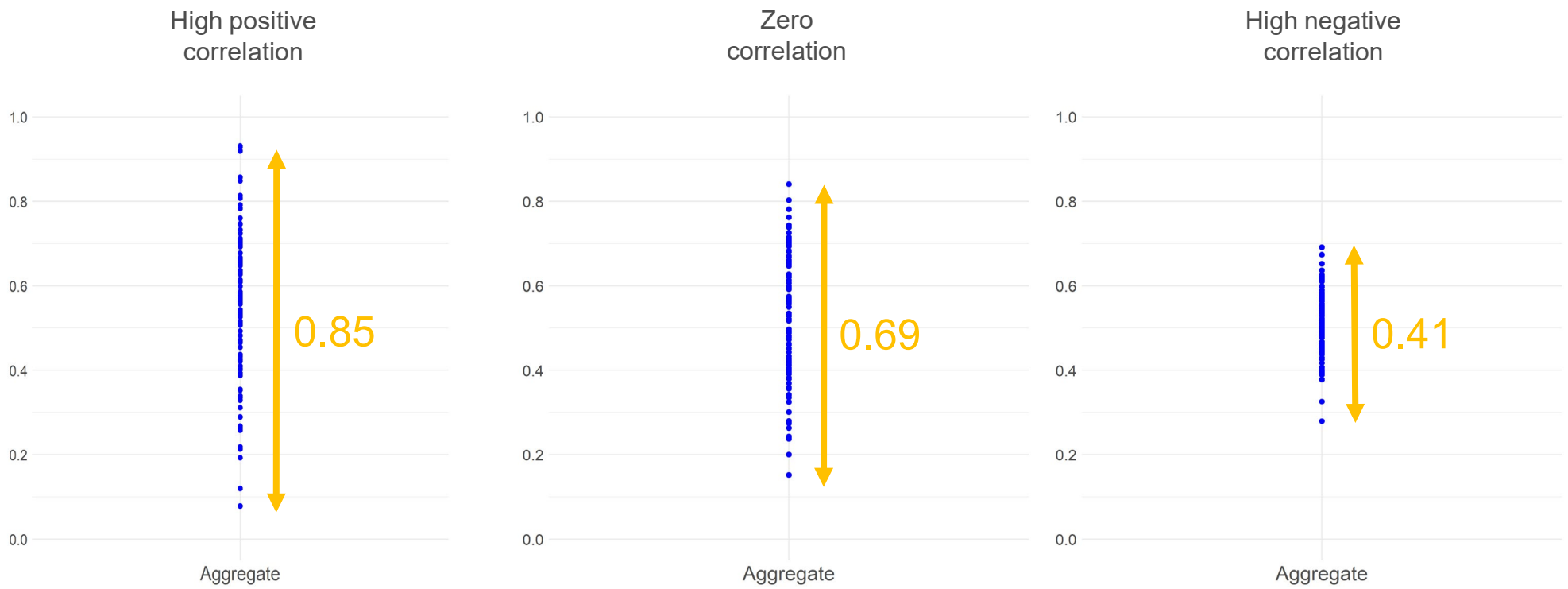
Correlation
-0.681





European Commission

Why correlations are important?



The range constitutes an implicit weight of the aggregates!