

The European Commission's science and knowledge service

Joint Research Centre



Principal Component Analysis and Reliability Analysis

Giulio Caperna and Hedvig Norlén

Outline

How multivariate methods can help to understand the statistical coherence of our composite indicator

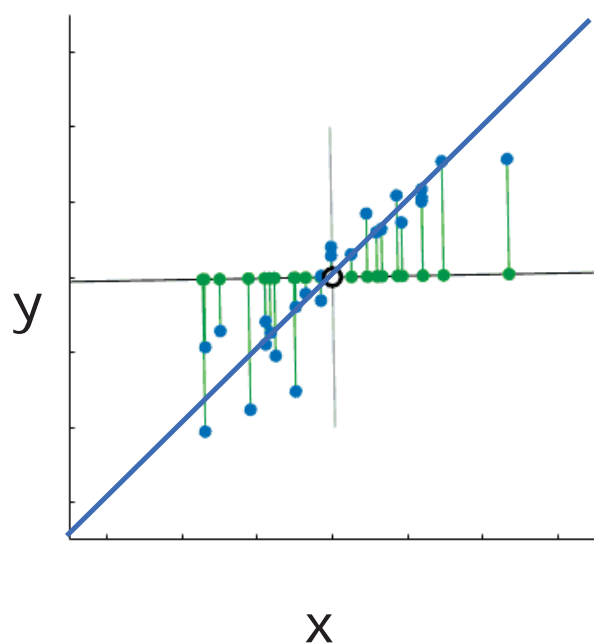
Part 1 Principal Component Analysis

Part 2 Reliability Analysis (Cronbach's alfa)

Is the structure of our composite indicator statistically well-defined?

Is the set of available indicators sufficient to describe the pillars and the sub-pillars of the composite indicator?

Principal Component Analysis (PCA)



We have data (more than one variable)

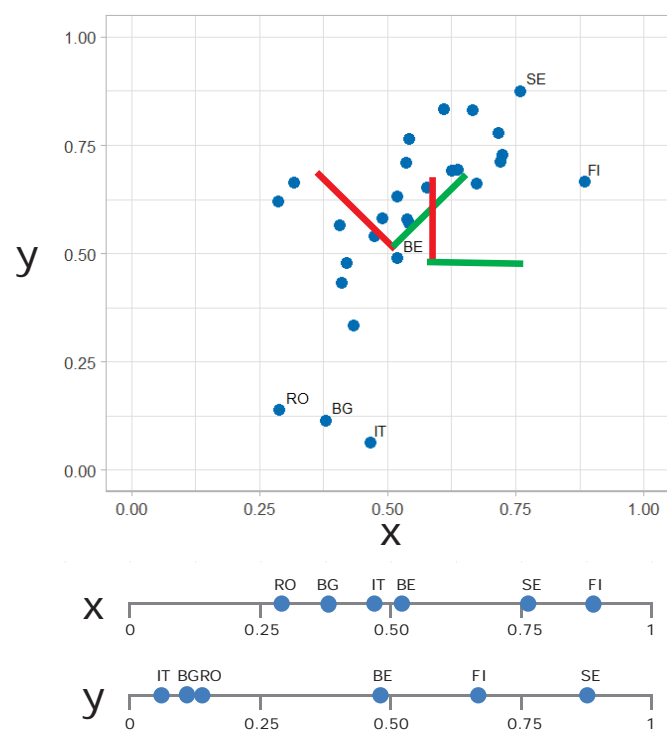
We want to describe the data with less variables!

PCA will find this “best” line:

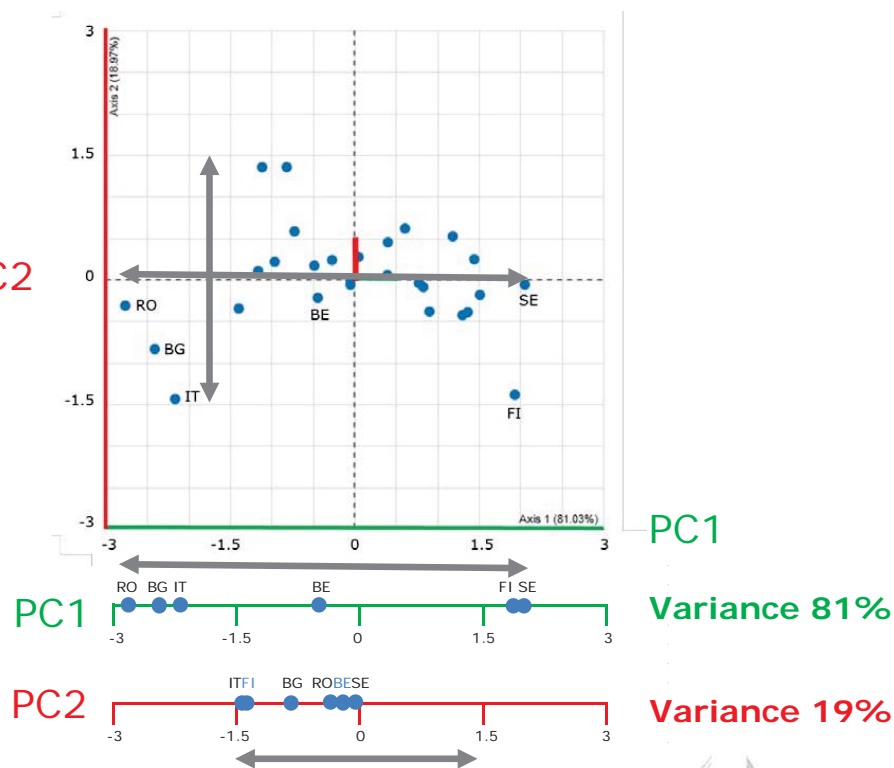
- 1) Maximum variance
- 2) Minimum error

First Principal Component!

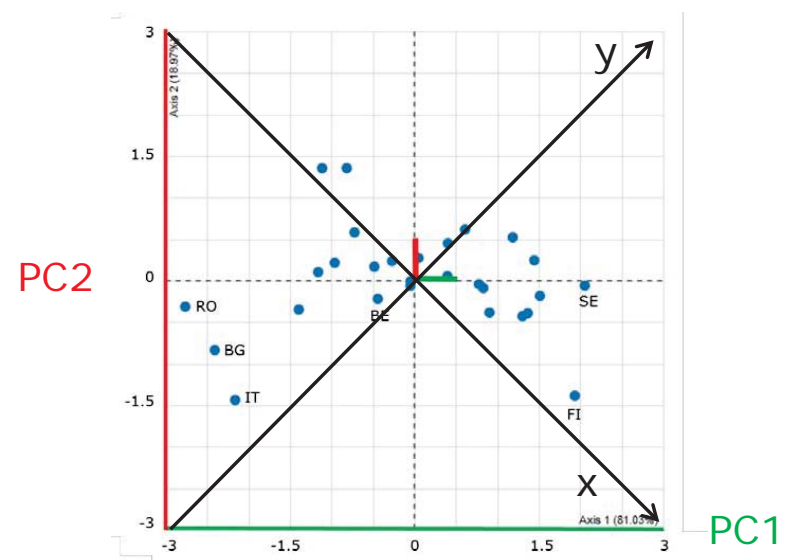
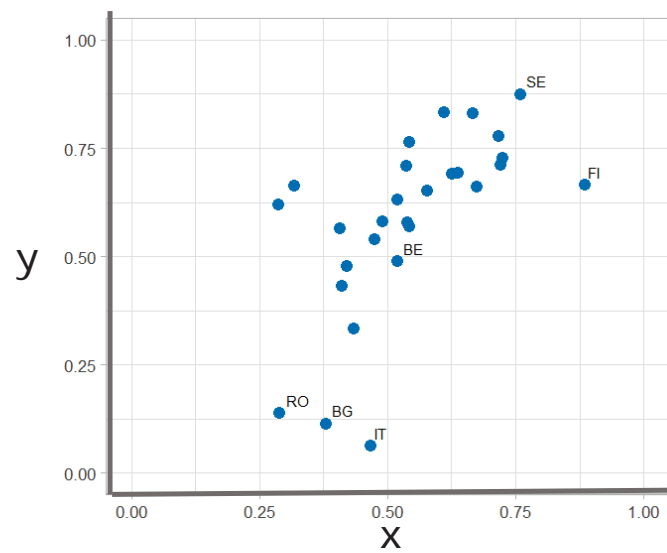
Example with two variables



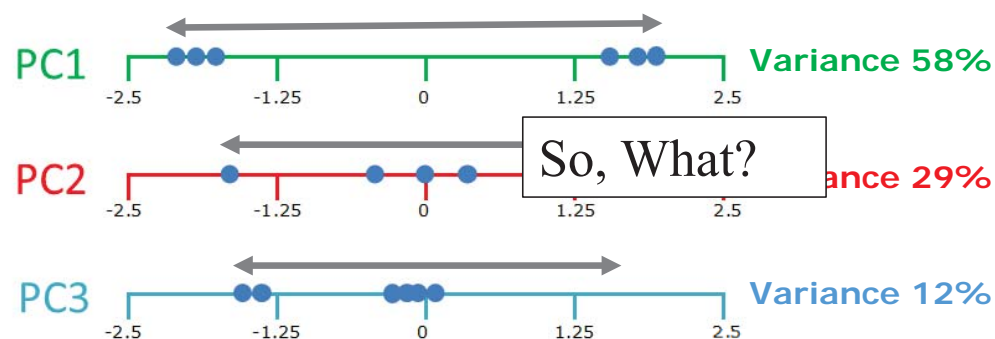
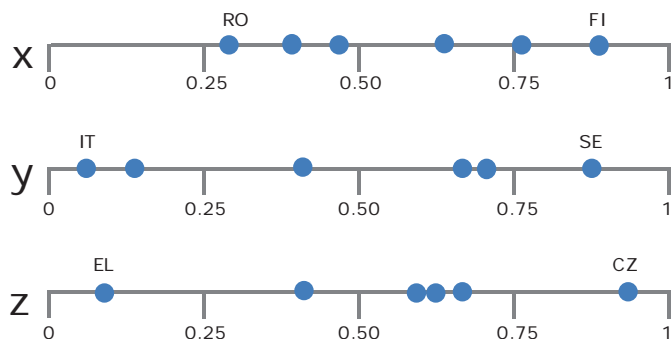
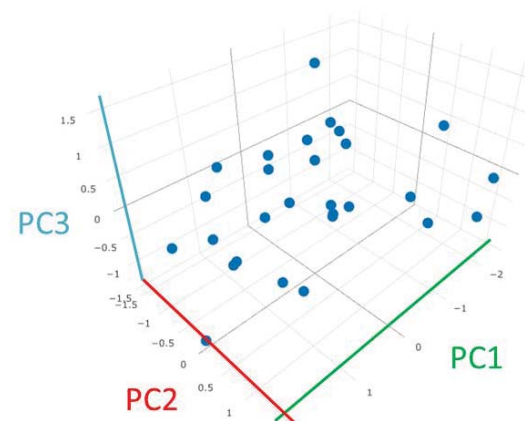
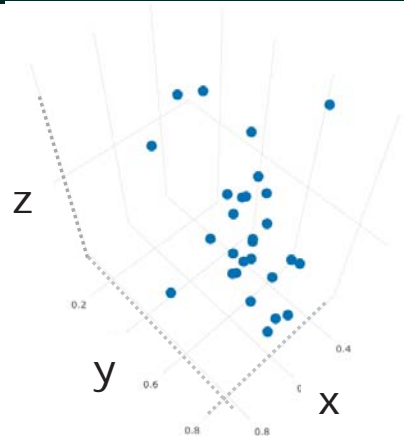
PC2



What about the original variables?



Example with three variables



How we use PCA - Introduction

A **Composite Indicator** measures multifaceted phenomenon - combination of different aspects (**Sub-pillars/Pillars**).

Each aspect can be measured by a set of observable variables (**Indicators**).

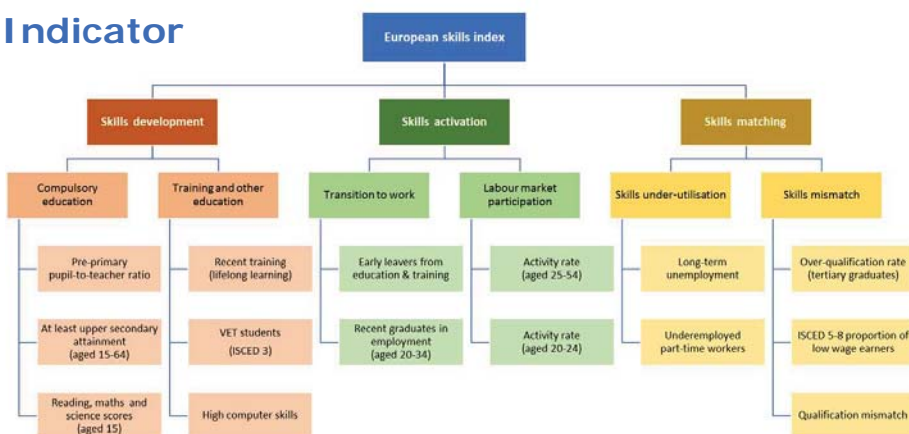
Framework of the European Skills Index 2018

Composite Indicator

Pillars

Sub-pillars

Indicators

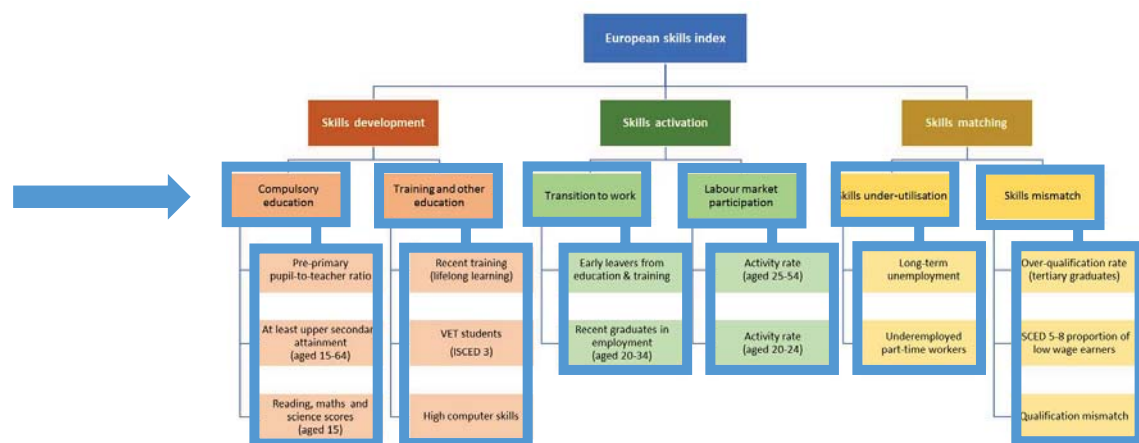


How we use PCA - unidimensionality

PCA is used to verify the internal consistency, verify **"unidimensionality"** within:

- 1) each Sub-pillar (across Indicators)

Framework of the European Skills Index 2018

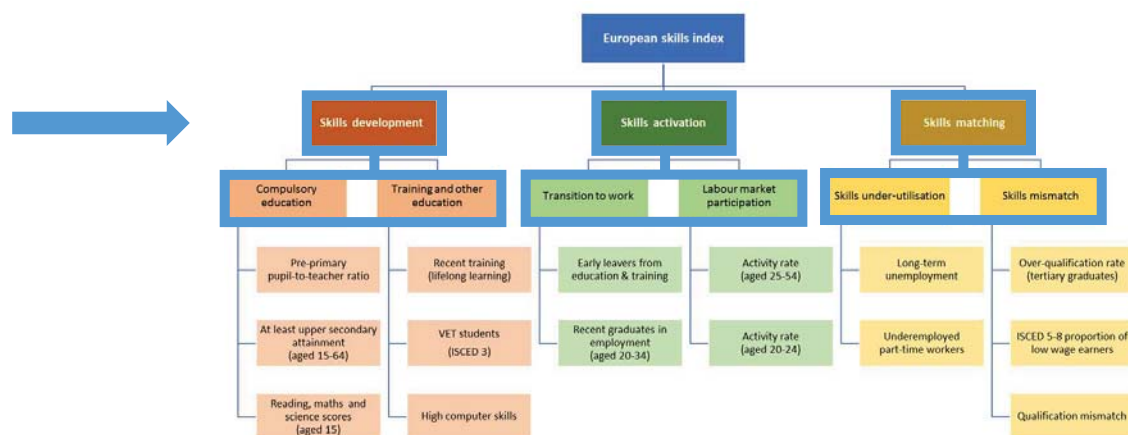


How we use PCA - unidimensionality

PCA is used to verify the internal consistency, verify **"unidimensionality"** within:

- 1) each Sub-pillar (across Indicators)
- 2) each Pillar (across Sub-pillars)

Framework of the European Skills Index 2018

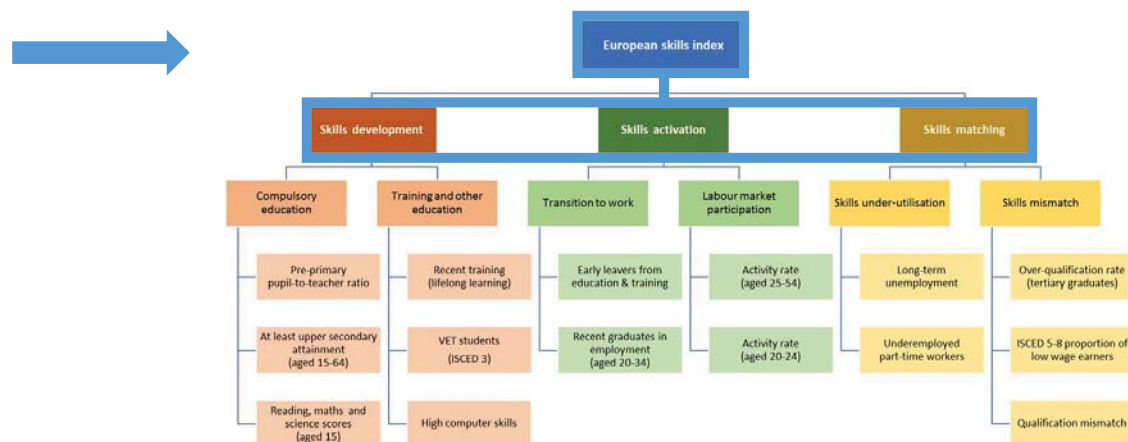


How we use PCA - unidimensionality

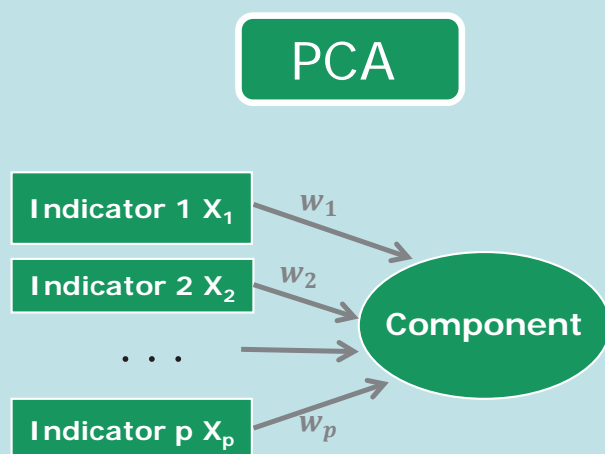
PCA is used to verify the internal consistency, verify **"unidimensionality"** within:

- 1) each Sub-pillar (across Indicators)
- 2) each Pillar (across Sub-pillars)
- 3) the Composite Indicator (across Pillars)

Framework of the European Skills Index 2018



PCA



Observed indicators are reduced into components

Identify a small number of PCs that explain most of the variance observed.

PCA summarizes information of all indicators and reduces it into a **fewer number of components**

Each principal component PC_i is a new variable computed as **a linear combination** of the original (standardized) variables

$$PC_1 = w_1x_1 + w_2x_2 + \cdots w_px_p$$

First steps in PCA

Check the **correlation structure** of the data and perform 2 “pre-tests”

1) Bartlett’s sphericity test

The test checks if the observed correlation matrix R diverges significantly from the identity matrix. $H_0 : |R| = 1$, $H_1 : |R| \neq 1$ (In English: «If we have good correlations»)

Want to reject H_0 to be able to do perform a valid PCA (Bartlett (1937))

2) Kaiser-Meyer-Olkin (KMO) Measure for Sampling Adequacy

Measure of the strength of relationship among variables based on correlations and partial correlations. KMO between $[0; 1]$.

Want KMO close to 1 to be able to perform a valid PCA. $KMO > 0.6$ OK!
(Kaiser (1970), Kaiser-Meyer (1974))

First steps in PCA

2) Kaiser-Meyer-Olkin (KMO) Measure for Sampling Adequacy

Kaiser's own interpretations of the KMO values

KMO values:

**"in the .90s, marvelous
in the .80s, meritorious
in the .70s, middling
in the .60s, mediocre
in the .50s, miserable
below .50, unacceptable"**

1) Bartlett's sphericity test and
2) KMO test provide info
whether it is possible to do PCA,
but do not give info of the
"magic number" - how many
components are needed

KMO > 0.6 OK

How to get the PCs

1) Eigenvalue decomposition of a data correlation matrix
(case of composite indicator),

2) Singular Value Decomposition (SVD) of a data matrix
after mean centering (normalizing) the data matrix

Important assumption:
Linearity of relations

Important property:
The PCs are orthogonal unless they are rotated

Finding the “magic number” - determining how many components in PCA

Several methods exist. The 3 most common are:

1) Kaiser–Guttman ‘Eigenvalues greater than one’ criterion

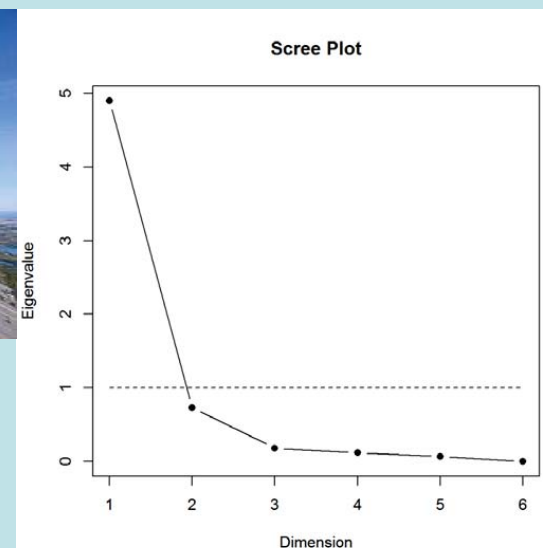
(Guttman (1954), Kaiser (1960)). Select all components with eigenvalues over 1 (or 0.9)

2) Cattell’s scree test

(Cattell (1966)) “Above the elbow” approach

3) Certain percentage of explained variance

e.g., $>2/3$, 75%, 80%,...



Social Progress Index

2018 SOCIAL
PROGRESS INDEX

SOCIAL
PROGRESS
IMPERATIVE

The **Social Progress Index (SPI)** is an international monitoring framework for measuring social progress without resorting to the use of economic indicators.

146 fully and 90 partially ranked countries.

2018 SOCIAL
PROGRESS INDEX
Methodology Summary



By Scott Stern, Amy
Wares & Tamar Epner

SOCIAL
PROGRESS
IMPERATIVE

2018 SOCIAL
PROGRESS INDEX
Executive Summary



SOCIAL
PROGRESS
IMPERATIVE

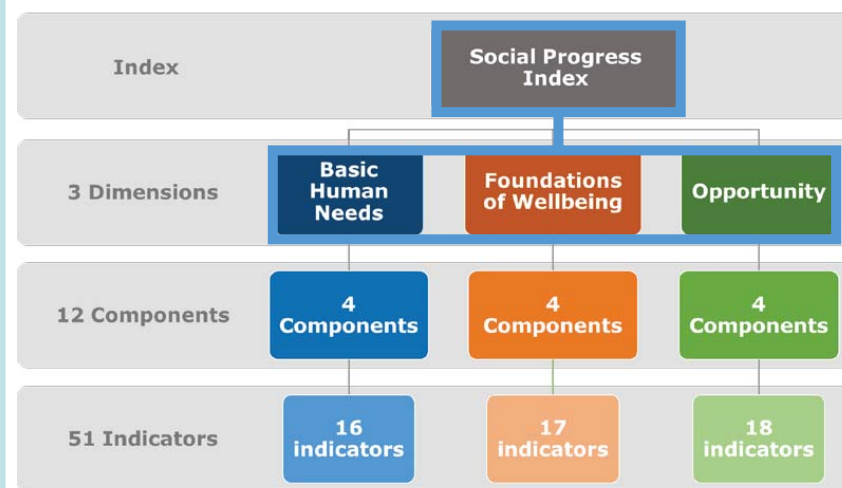
Social Progress Index - PCA at Index level

Social progress is measured taking into account the following **three broad aspects**:

1. Meeting everyone's basic needs for food, clean water, shelter and security.
2. Living long, healthy lives with basic knowledge and communication and a clean environment.
3. Practicing equal rights and freedoms and pursuing higher education.

5th version of SPI was launched September 2018.

Let's try PCA



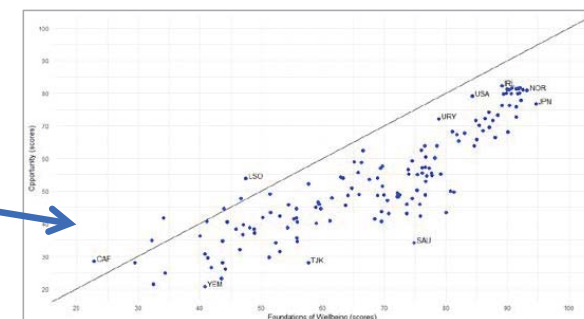
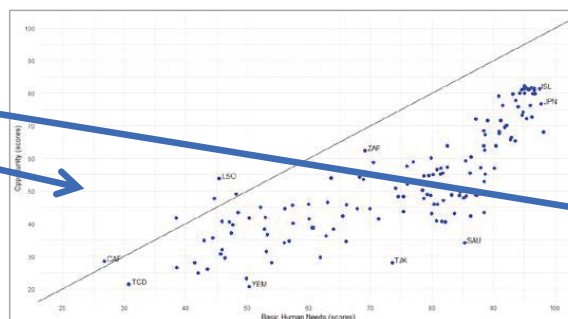
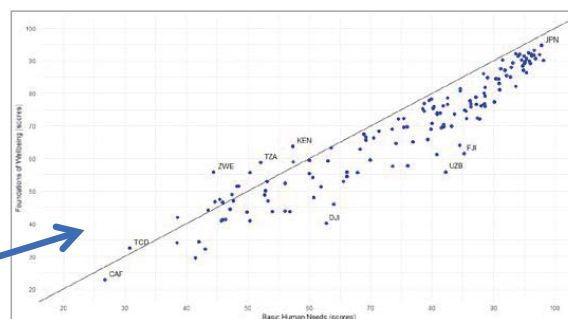
Social Progress Index - PCA example

Check the correlation structure

Correlation matrix	Basic Human Needs	Foundations of Wellbeing	Opportunity
Basic Human Needs	1		
Foundations of Wellbeing	0.95	1	
Opportunity	0.81	0.88	1

Pearson correlation coefficients

(Significance level $\alpha = 0.01$, $n=146$, critical value = 0.21)



Social Progress Index - PCA example

Check the correlation structure

Correlation matrix	Basic Human Needs	Foundations of Wellbeing	Opportunity
Basic Human Needs	1		
Foundations of Wellbeing	0.95	1	
Opportunity	0.81	0.88	1

Pearson correlation coefficients

1) Bartlett's sphericity test

p-value ($2.0e-116$) < 0.01 **Reject H_0**

2) KMO Measure for Sampling Adequacy

Overall MSA = 0.69 **KMO > 0.6**

Social Progress Index - PCA results

$$2.76 = 0.96^2 + 0.98^2 + 0.93^2$$

Total variance explained

$$92.02 = 2.76 / 3 * 100$$

Component	Eigenvalue	Variance	Cumulative variance
1	2.76	92.02	92.02
2	0.20	6.55	98.57
3	0.04	1.43	100,00
Sum	3	100	

Stopping criterion Eigenvalue > 1

$$PC_1 = 0.96X_{BHN} + 0.98X_{FoW} + 0.93X_{Opp}$$

Pearson correlation coefficients between pillar and principal component

	Component loadings		
	PC1	PC2	PC3
1. Basic Human Needs	0.96	-0.25	0.12
2. Foundations of Wellbeing	0.98	-0.09	-0.16
3. Opportunity	0.93	0.35	0.05
Sum of Squares	2.76	0.20	0.04

One dimension verified!

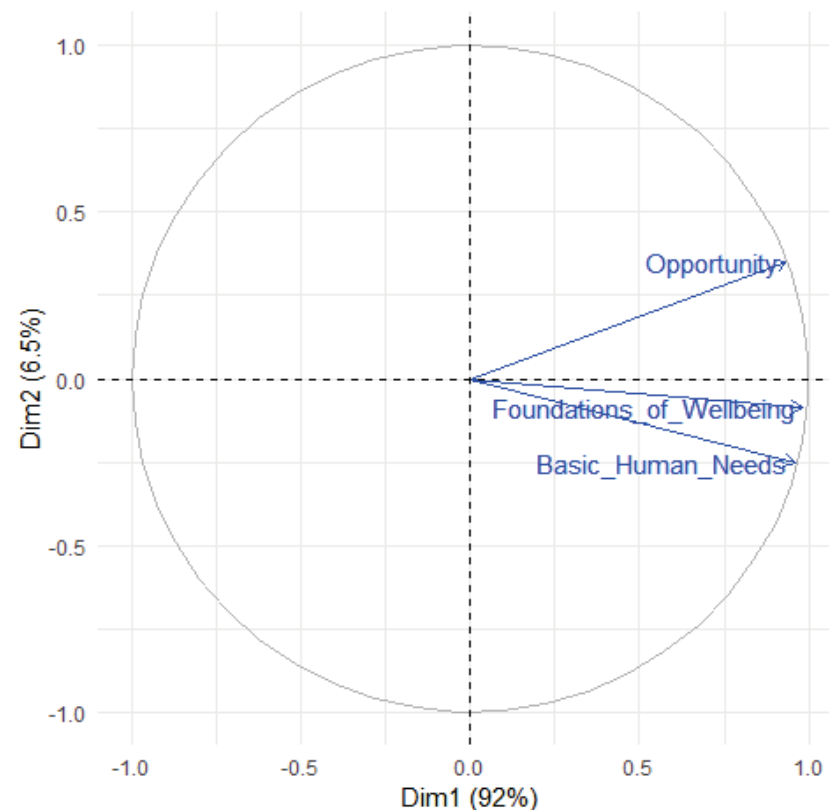
Social Progress Index - PCA results

PCA results often summarized in a "Factor map"

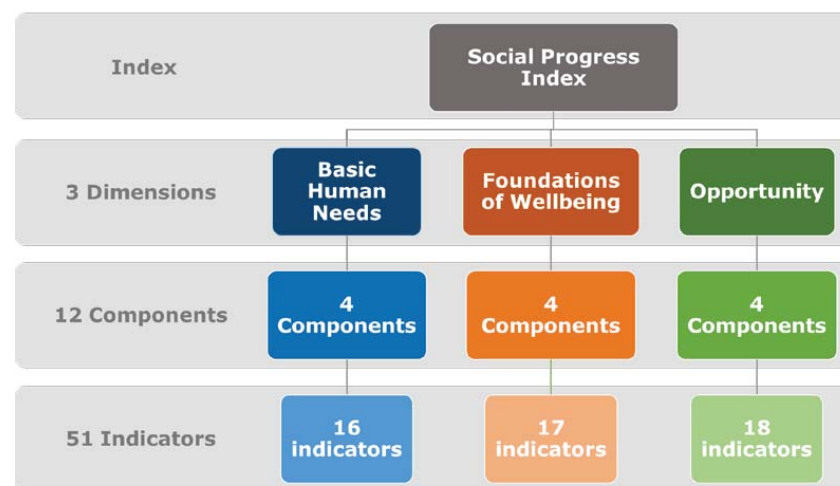
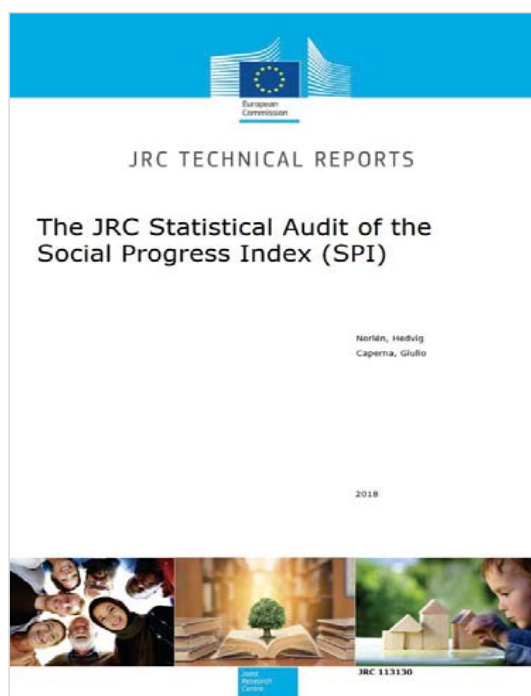
$$PC_1 = 0.96X_{BHN} + 0.98X_{FoW} + 0.93X_{Opp}$$

$$PC_2 = -0.25X_{BHN} - 0.09X_{FoW} + 0.35X_{Opp}$$

PC_2 Second principal component is useful to evaluate the differences between the first two and third dimensions.



Social Progress Index - Audit results



Social Progress  @socprogress

Thanks @EU_ScienceHub for auditing the 2018 #SocialProgress Index, showing it is a rigorous measure of social outcomes and useful catalyst for change #data #EUknowledge ow.ly/hkmh30lZYVp

Social Progress Index - Audit results

Redundancies in SPI framework - very strong correlations between the SPI aggregates (dimensions and components).

PCA was also performed on the whole set of 51 indicators

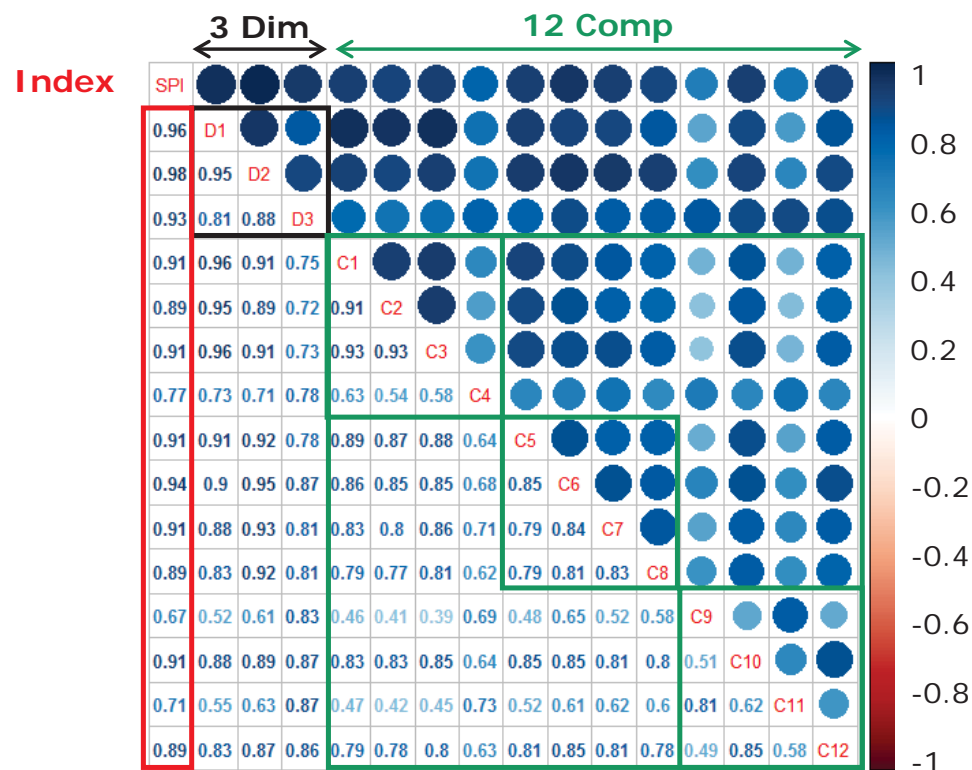
Some components measure similar concepts:

C1: Water and Sanitation

C2: Shelter

C5: Access to Basic Knowledge

C12: Access to Advanced Education



Social Progress Index - Audit results

Very strong correlations between the SPI aggregates (dimensions and components).

PCA was also performed on the whole set of 51 indicators.

7 latent dimensions (principal components) are retrieved, which capture 78% of the total variance in the underlying indicators.

Total variance explained

Component	Eigenvalue	Cumulative variance
1	27,55	54,02
2	5,30	64,42
3	1,98	68,31
4	1,63	71,51
5	1,24	73,94
6	1,22	76,34
7	1,07	78,44
8	0,97	80,33
9	0,84	81,98
10	0,76	83,46
11	0,67	84,78
12	0,67	86,10
...		
51	0,01	100,00
Sum	51	

Difference less than 8%

Stopping criterion Eigenvalue > 1

How do we use the results?

First a question:

What do we say to the “all-knowing” statistical method?

Let me check the Conceptual Framework!
(and steps 2 to 6)

Social Progress Index - Audit results

From JRC Audit report:

"Less number of components is also in line with the results from the "beyond GDP" approach by the Commission on the Measurement of Economic Performance and Social Progress. The SPI conceptual framework has been influenced by this work. **So apart from the statistical reasoning, the result from the "beyond GDP" approach, may also give conceptual justification for reducing the number of components in the SPI framework.**"*

* Stiglitz, J., Sen, A., Fitoussi, J.-P., 2009. Report by the Commission on the Measurement of Economic Performance and Social Progress. In: Commission on the Measurement of Economic Performance and Social Progress, Paris, France.

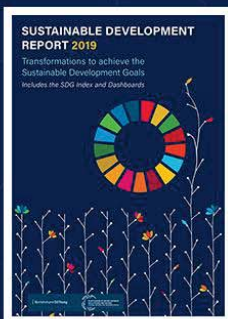
An example of Transparency

SUSTAINABLE DEVELOPMENT REPORT 2019

Transformations to achieve the
Sustainable Development Goals
Includes the SDG Index and Dashboards



SDG	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	Index
1	1.00																	
2	0.49	1.00																
3	0.84	0.64	1.00															
4	0.77	0.61	0.84	1.00														
5	0.35	0.54	0.59	0.61	1.00													
6	0.69	0.66	0.81	0.73	0.68	1.00												
7	0.88	0.50	0.85	0.81	0.46	0.71	1.00											
8	0.50	0.60	0.68	0.62	0.59	0.63	0.51	1.00										
9	0.66	0.66	0.82	0.67	0.59	0.75	0.68	0.63	1.00									
10	0.36	0.29	0.36	0.17	0.01	0.21	0.19	0.24	0.38	1.00								
11	0.54	0.46	0.68	0.66	0.64	0.69	0.59	0.53	0.55	0.10	1.00							
12	-0.59	-0.53	-0.76	-0.59	-0.52	-0.67	-0.60	-0.46	-0.86	-0.32	-0.50	1.00						
13	-0.28	-0.17	-0.33	-0.32	-0.18	-0.19	-0.27	-0.10	-0.32	-0.05	-0.13	0.50	1.00					
14	-0.17	-0.01	-0.11	-0.10	0.09	-0.04	-0.13	0.06	-0.05	-0.19	-0.04	0.04	0.02	1.00				
15	-0.09	0.15	-0.02	-0.03	0.17	0.12	-0.05	0.14	0.09	0.01	-0.01	-0.02	0.22	0.24	1.00			
16	0.64	0.59	0.81	0.67	0.50	0.65	0.60	0.58	0.78	0.43	0.63	-0.72	-0.33	-0.11	0.04	1.00		
17	0.19	0.07	0.21	0.22	0.14	0.12	0.29	-0.03	0.11	0.01	0.16	-0.12	-0.35	-0.02	-0.08	0.14	1.00	
Index	0.84	0.71	0.93	0.86	0.67	0.86	0.86	0.73	0.83	0.40	0.73	-0.68	-0.20	-0.01	0.14	0.79	0.24	1.00



Sustainable Development Report 2019

Transformations to Achieve the Sustainable Development Goals

The Sustainable Development Report 2019 presents the SDG Index and Dashboards for all UN member states and frames the implementation of the Sustainable Development Goals (SDGs) in terms of six broad transformations. It was prepared by teams of independent experts at the Sustainable Development Solutions Network (SDSN) and the Bertelsmann Stiftung.

[Download the report](#)

Related News & Media

[New Report Shows Major Long-Term Transformations are Necessary for Achieving the Sustainable Development Goals](#)

[Australia ranked 38th globally on sustainable development](#)

[Canada ranked 20th globally on sustainable development](#)

[España ocupa el 21º puesto en desarrollo sostenible a nivel mundial](#)

[La France au 4ème rang mondial en matière de développement durable](#)

[The United States ranked 35th globally on sustainable development](#)

Explore the data



Interactive SDG Dashboards

The interactive 2019 SDG Dashboards provide a visual representation of countries' performance by SDGs to identify priorities for action.

Additional Materials

[2019 G20 and Large Countries Edition](#)

[Methodology Paper \(2018\)](#)

[Executive Summary \(Adobe Spark\)](#)

[SDR2019 – Transformations to achieve the SDGs \(Prezi\)](#)

[About the Authors](#)

[Citation](#)

[JRC Statistical Audit](#)

[Behind the numbers: Joint Research Centre Audit of the SDG Index and Dashboards](#)

Download the data

[2019 Results \(Excel\)](#)

[2019 Results \(STATA\)](#)

[2019 Country Profiles](#)

[View on Github](#)



JRC TECHNICAL REPORTS

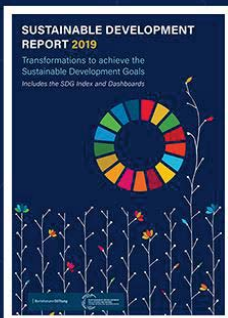
JRC Statistical Audit of the Sustainable Development Goals Index and Dashboards

Papadimitriou, Eleni
Neves, Ana Rita
Becker, William

2019



European
Commission



Sustainable Development Report 2019

Transformations to Achieve the Sustainable Development Goals

The Sustainable Development Report 2019 presents the SDG Index and Dashboards for all UN member states and frames the implementation of the Sustainable Development Goals (SDGs) in terms of six broad transformations. It was prepared by teams of independent experts at the Sustainable Development Solutions Network (SDSN) and the Bertelsmann Stiftung.

[Download the report](#)

Related News & Media

[New Report Shows Major Long-Term Transformations are Necessary for Achieving the Sustainable Development Goals](#)

[Australia ranked 38th globally on sustainable development](#)

[Canada ranked 20th globally on sustainable development](#)

[España ocupa el 21º puesto en desarrollo sostenible a nivel mundial](#)

[La France au 4ème rang mondial en matière de développement durable](#)

[The United States ranked 35th globally on sustainable development](#)

BEHIND THE NUMBERS: JOINT RESEARCH CENTRE AUDIT OF THE SDG INDEX AND DASHBOARDS

May 30, 2019

By Eleni Papadimitriou, Ana Rita Neves and William Becker, European Commission's Joint Research Centre

Indexes are powerful policy, advocacy and communication tools, but such as sustainable development cannot synthesize large and complex data sets in an effective means to stimulate public debate.

However, the process of building an index faces methodological challenges. In order to have transparency of their indexes, the European Indicators and Scoreboards (COIN) at the audits of indexes upon request of internal services. To date, over 100 indexes, in a variety of formats, have been recommended from our research team.

Additional Materials

[2019 G20 and Large Countries Edition](#)

[Methodology Paper \(2018\)](#)

[Executive Summary \(Adobe Spark\)](#)

[SDR2019 – Transformations to achieve the SDGs \(Prezi\)](#)

[About the Authors](#)

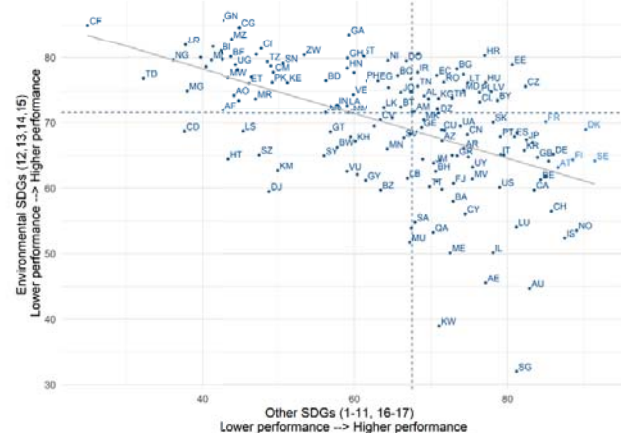
[Citation](#)

[JRC Statistical Audit](#)

[Behind the numbers: Joint Research Centre Audit of the SDG Index and Dashboards](#)

Plotting the environmental SDGs (12, 13, 14 and 15), which present a negative or insignificant association with the overall index against all the other SDGs, shows that the top five countries of the SDG Index are actually ranked well below many other countries with lower performance on the SDG Index. Denmark, which tops the list on the SDG Index, is in 143rd position for SDG12. In the other direction, the Central African Republic, which is at the bottom of the SDG Index, holds the second best position on SDG13.

Relation between the environmental SDGs (12,13,14,15) and the other SDGs



The SDG Index is a noteworthy and much-needed tool which synthesizes the 17 adopted SDGs into a single measure. Overall, the ranks of the SDG Index are fairly robust. The index rigorously follows

Part 2: Reliability Analysis

Cronbach's α is measure of internal consistency and reliability

Based upon classical test theory. Most common estimate of reliability, Cronbach (1951).

Cronbach's α is defined as Ratio of the Total **Covariance** in the "test" to the Total **Variance** in the "test".

By "test" we mean the set of indicators constituting a Sub-pillar/Pillar/Composite Indicator.

Cronbach's α

Cronbach's α is measure of internal consistency and reliability

α ranges from 0 to 1. If all indicators are entirely independent from one another (*i.e.* are not correlated or share no covariance) $\alpha = 0$. If all of the items have high covariances, then α will approach 1.

Caution! α increases when the number of indicators increases.

Cronbach's α cut off values

Cronbach's α is measure of internal consistency and reliability



Lee Cronbach
(1916-2001)

Cronbach's alpha	Internal consistency
$0.9 \leq \alpha$	Excellent
$0.8 \leq \alpha < 0.9$	Good
$0.7 \leq \alpha < 0.8$	Acceptable
$0.6 \leq \alpha < 0.7$	Questionable
$0.5 \leq \alpha < 0.6$	Poor
$\alpha < 0.5$	Unacceptable

Cronbach's α - Social Progress Index

Cronbach's α is measure of internal consistency and reliability

Social Progress Index	Cronbach's Alpha	
Index level (3 dimensions)	0.96	
Item level	Corrected Item-Total Correlation	Alpha if Item is dropped
1: Basic Human Needs	0.91	0.94
2: Foundations of Wellbeing	0.96	0.89
3: Opportunity	0.86	0.97

Cronbach's alpha

$0.9 \leq \alpha$

$0.8 \leq \alpha < 0.9$

$0.7 \leq \alpha < 0.8$

$0.6 \leq \alpha < 0.7$

$0.5 \leq \alpha < 0.6$

$\alpha < 0.5$

Internal consistency

Excellent

Good

Acceptable

Questionable

Poor

Unacceptable

If α is too high it may suggest that some items are redundant as that are measuring very similar concepts. **Maximum α 0.9 has been recommended.**

Cronbach's α – SDG Index 2019

SDG Index	Cronbach's Alpha
Index level (17 goals)	0.62
Removing Goals 12-15	0.81

Cronbach's alpha

$0.9 \leq \alpha$

$0.8 \leq \alpha < 0.9$

$0.7 \leq \alpha < 0.8$

$0.6 \leq \alpha < 0.7$

$0.5 \leq \alpha < 0.6$

$\alpha < 0.5$

Internal consistency

Excellent

Good

Acceptable

Questionable

Poor

Unacceptable



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/>

The European Commission's
Competence Centre on Composite
Indicators and Scoreboards

