

# **Semestrální projekt 2**

**ENC-PSMM**

**Analýza hlavních komponent  
a faktorová analýza**

**Bc. Natalie Drábková**

**Brno 2025**



# Obsah

<b>1</b>	<b>Index sociálního pokroku.....</b>	<b>4</b>
1.1	Úvod.....	4
1.2	Data.....	4
<b>2</b>	<b>Analýza hlavních komponent.....</b>	<b>6</b>
2.1	Vyhodnocení skóre hlavních komponent .....	6
2.2	Sutinový graf .....	7
<b>3</b>	<b>Faktorová analýza .....</b>	<b>8</b>
3.1	Faktorová analýza s rotací varimax (2 faktory) .....	8
3.2	Faktorová analýza s rotací varimax (3 faktory) .....	9
<b>4</b>	<b>Závěr .....</b>	<b>11</b>
	<b>Přílohy.....</b>	<b>12</b>

# 1 Index sociálního pokroku

## 1.1 Úvod

Téma projektu se týká indexu sociálního pokroku (SPI), který je založen na třech hlavních dimenzích: *Základní potřeby*, *Základy blahobytu a Příležitosti*. Každá dimenze má čtyři kategorie, jež každá obsahuje několik ukazatelů. SPI index celkem zahrnuje 50 indikátorů.

Cílem projektu je provést analýzu hlavních komponent, která poslouží k určení vhodného počtu faktorů a následné provedení faktorové analýzy.

## 1.2 Data

Ze všech 50 proměnných bylo kvůli chybějícím hodnotám deset ukazatelů odstraněno. Zároveň byl odstraněn stát Fiji, který i po redukci proměnných měl jednu chybějící proměnnou. Aby nemuselo dojít k odstranění celého ukazatele, byla tato země redukována s vědomím, že nebude mít příliš velký dopad na danou analýzu.

V příloze tohoto dokumentu je výpis 40 proměnných, které byly použity v této práci. Data zahrnovaly údaje pro 162 států a sledovaným obdobím byl rok 2016.

Před samotnou analýzou byla nejprve uskutečněna standardizace dat, aby byly na výstupu korelace.

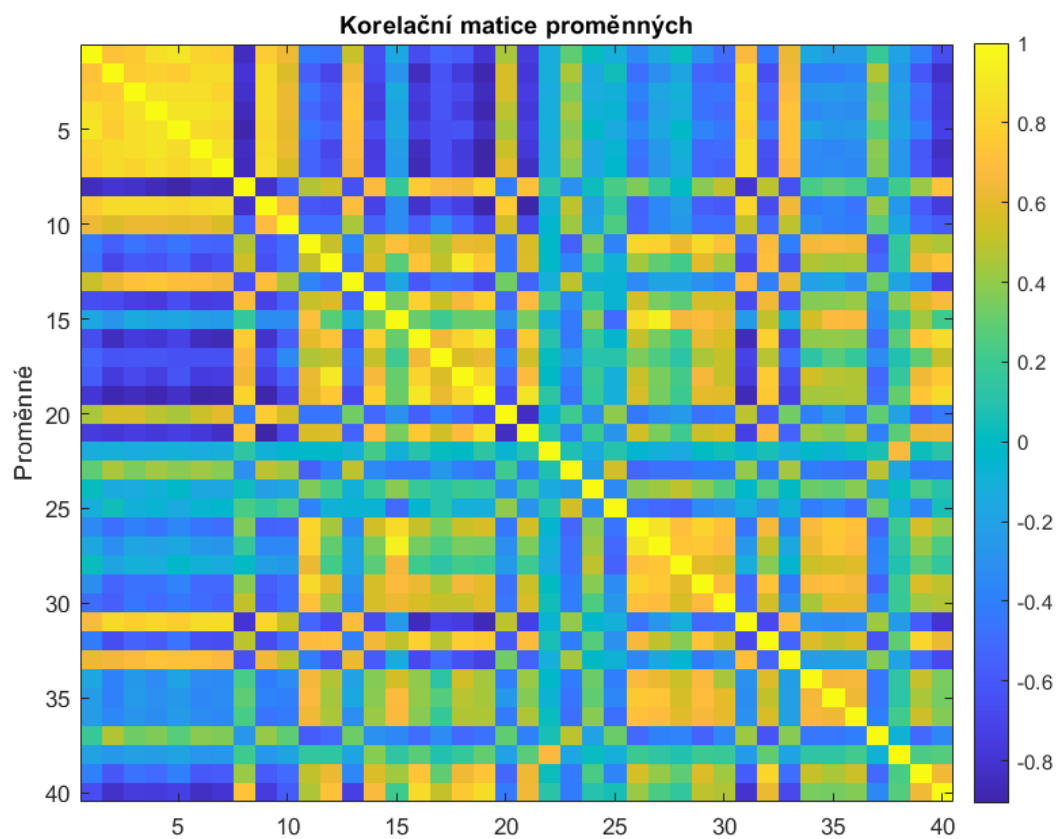
Byla provedena detekce odlehlých hodnot, kde se jako nejextrémnější jeví Čína s hodnotou  $T2 = 151.22$  a Spojené státy americké s hodnotou  $T2 = 127.29$ . I přes identifikaci několika dalších odlehlých států (Severní Korea, Kuba, Středoafriická republika), byly tyto země spolu s Čínou a USA ponechány v analýze. Důvodem je jejich významná výpovědní hodnota o extrémech v rámci sociálního pokroku. Odlehlost zde odráží reálnou rozmanitost světových socioekonomických podmínek.

Na Obr. 1 lze vidět graf tepelné mapy, který slouží k ověření, zda jsou v datech přítomny vzájemné vztahy mezi proměnnými, což je důležitým předpokladem pro použití metody PCA<sup>1</sup> a faktorové analýzy. Lze vidět, že některé proměnné jsou vzájemně silně korelované, neboť se koeficient blíží 1 či -1. Například mezi přístupem ke kvalitní zdravotní péči a dětskou úmrtností byla zjištěna silná negativní korelace ( $r = -0,736$ ). Tento vztah je očekávaný – čím lepší je dostupnost kvalitní lékařské péče, tím nižší bývá úmrtnost dětí. To potvrzuje vhodnost použití PCA a faktorové analýzy.

Analýza hlavních komponent i faktorová analýza byly vypracovány v software MATLAB.

---

<sup>1</sup> PCA je označení pro analýzu hlavních komponent (*Principal Component Analysis*)



Obr. 1: Heatmapa korelačních koeficientů proměnných

## 2 Analýza hlavních komponent

Pro 162 států byla provedena metoda hlavních komponent v rámci SPI ukazatelů.

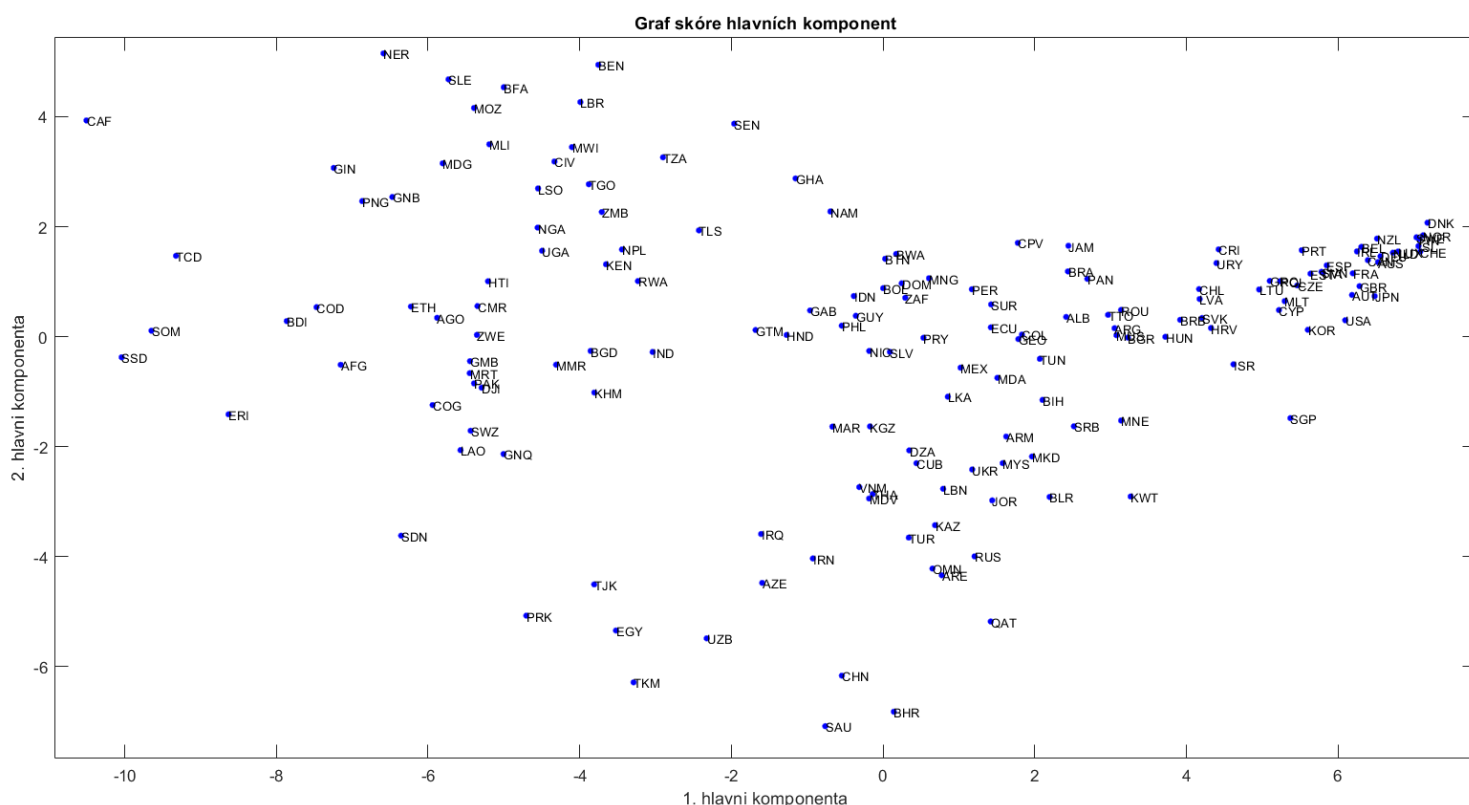
### 2.1 Vyhodnocení skóre hlavních komponent

Obr. 2 prezentuje graf skóre pro dvě hlavní komponenty (PC1 a PC2), aplikované na ukazatele tvořící index sociálního pokroku (SPI). Každý bod reprezentuje skóre pro daný stát.

První hlavní komponenta vysvětlila největší část variability mezi zeměmi, a to konkrétně 51 %. Na základě rozmístění států v rámci PC1 lze předpokládat, že tato komponenta souvisí s celkovou úrovní sociálního rozvoje, neboť na pravé straně grafu se seskupují země s vysokým skóre ukazatelů SPI. Jedná se především o Skandinávské země, Švýcarsko, Velkou Británii, Kanadu, Austrálii, Nový Zéland, Japonsko, Jižní Koreu a velký podíl států EU. To značí, že tyto státy dosahují nadprůměrných výsledků ve většině SPI indikátorech. Naproti tomu se v levé části grafu soustředí státy s velmi nízkými hodnotami jako Somálsko, Čad a Středoafriická republika.

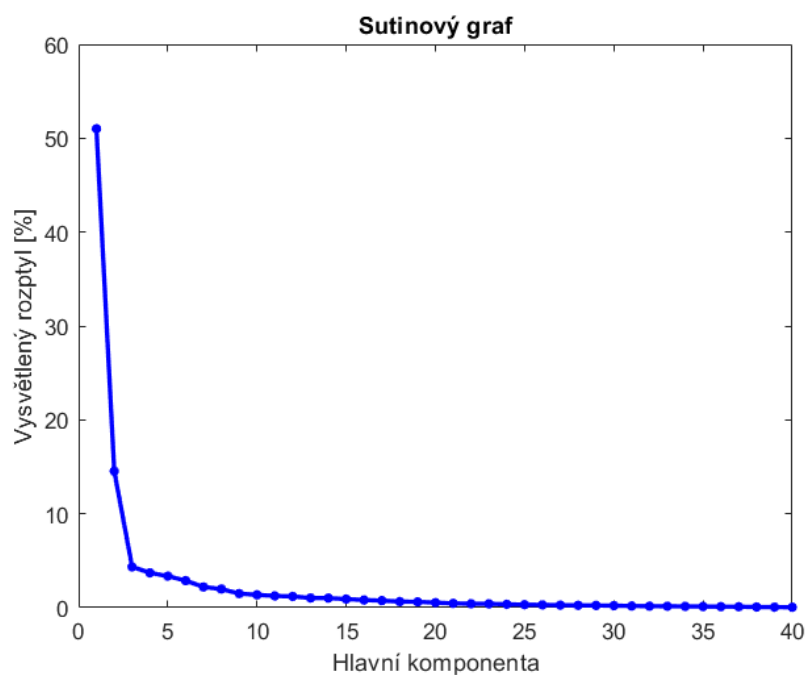
Druhá komponenta vysvětlila 15 % variability a rozlišuje spíše mezi zeměmi s odlišnými vývojem sociálního rozvoje. Státy, které mohou mít lepší výsledky v jedné dimenzi SPI, mohou pak zaostávat v jiné. Například Niger či Guinea mají vyšší hodnoty PC2, což může naznačovat silnější výkon v určité oblasti SPI, ale celkově nízké skóre.

Lze vidět, že vpravo se shromažďují spíše ekonomicky vyspělé demokratické země, mezi nimiž nejsou výrazné rozdíly v hlavních ukazatelích SPI. Naopak vlevo (a spíše výše) se nachází rozvojové země, zejména z Afriky, s velmi nízkým celkovým skóre SPI. V dolní části uprostřed mají státy jako Čína, Saudská Arábie, Turkmenistán či Írán nízké skóre v rámci druhé komponenty (PC2), což může být dáno problémy těchto států v oblastech osobní svobody a lidských práv, svobody tisku, (ne)demokracie nebo také participace občanů na rozhodování. Ačkoliv se svým „rozvojem“ (PC1) blíží vyspělým zemím, strukturálně jsou velmi odlišné ve svém sociálně-politickém uspořádání (PC2). Skupina nacházející se od středu mírně vpravo zahrnuje některé země východní Evropy, Latinské Ameriky, jihovýchodní Asie nebo severní Afriky. Jedná se o země se střední úrovní sociálního rozvoje jako je např. Bolívie, Indonésie, Maroko či Srbsko.



## 2.2 Sutinový graf

V rámci analýzy hlavních komponent byl vykreslen sutinový graf (Obr. 3), který zobrazuje vysvětlený rozptyl jednotlivými komponentami. Lze pozorovat výrazný zlom v druhé komponentě. Dochází také k výraznějšímu, avšak menšímu zlomu u třetí komponenty, nadále pak přínos k celkové variabilitě pozvolně klesá.



## 3 Faktorová analýza

Na základě výstupů analýzy hlavních komponent byly identifikovány významné zlomy u faktorů 2 a 3. Proto byly obě možnosti v rámci faktorové analýzy porovnány. Faktorová analýza byla provedena s rotací varimax, neboť umožňuje jasně přiřadit skupiny proměnných k faktorům a s tím související lepší interpretaci. Byly vyzkoušeny také rotace quarimax a bez rotace, avšak nedošlo ke srozumitelnější identifikaci daných faktorů, a tak se rotace varimax jevila jako nejvhodnější.

### 3.1 Faktorová analýza s rotací varimax (2 faktory)

Na základě faktorové analýzy s rotací varimax byly identifikovány dva hlavní faktory, které vysvětlují většinu variability mezi proměnnými.

Na Obr. 4 je znázorněn biplot, neboli graf faktorových zátěží. Lze vidět, že proměnné blíží se na ose x k hodnotě -1 či 1 spadají do faktoru 1. Naopak vektory proměnných blíží se na ose y k hodnotě -1 a 1 spadají do faktoru 2.

Faktor 1 vykazuje silné negativní zátěže u proměnných jako je infekční úmrtnost, úmrtnost matek, úmrtí v dopravě, dětská úmrtnost, úmrtí související s nebezpečnou vodou, sanitací a hygienou, populace využívající nekvalitní zdroje vody či hygienu, což značí velmi silný vliv těchto indikátorů na daný faktor. Silné pozitivní zátěže vykazují proměnné přístup k elektřině, ženy s vysokoškolským vzděláním, online správa, kvalita vzdělání, uživatelé internetu, přístup k základním službám, délka života, a zdravotní péče. Proměnné související s tímto faktorem především vyjadřují přístup ke kvalitní zdravotní péči a výživě, délku života a úmrtnost, dostupnost pitné vody a hygienických zařízení a elektrické připojení. Faktor 1 lze tak identifikovat jako Základní lidské potřeby a kvalita života.

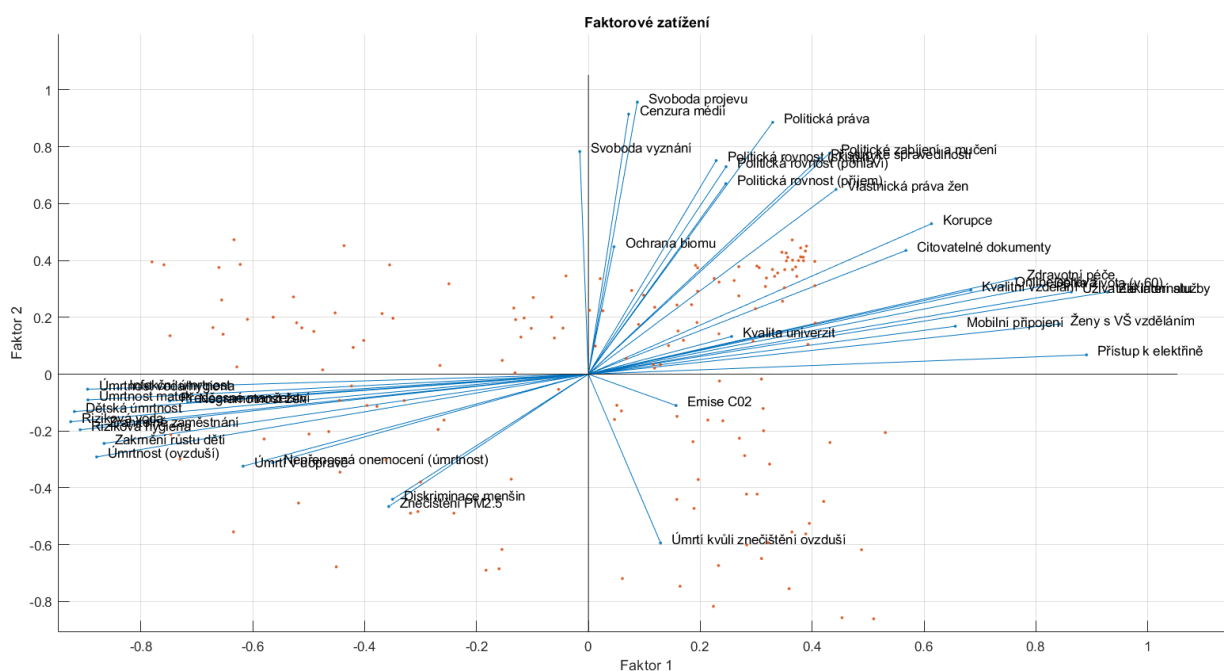
Faktor 2 vykazuje silné pozitivní zátěže především u proměnných jako svoboda projevu, cenzura médií, politická práva, svoboda vyznání, politické zabíjení a mučení, vlastnická práva žen, korupce, rovnost politické moci podle sociálních skupin a pohlaví. Tento faktor popisuje proměnné týkající se politické a osobní svobody, genderové rovnosti a přístupu k informacím. Faktor 2 je možné interpretovat jako Osobní svobody, občanská práva a příležitosti participace.

Z grafu faktorových zátěží lze pozorovat, že proměnné patřící do faktoru 1 jsou nekorelované s proměnnými faktoru 2, neboť mezi sebou svírají úhel 90°. Například ukazatel svobody vyznání a úmrtnost matek nevykazují žádnou závislost. Naopak proměnné, které se téměř překrývají, spolu korelují – například svoboda projevu a cenzura médií.

Proměnné, jejichž zátěže vykazovala hodnotu pod  $\pm 0.5$  u obou faktorů, byly považovány za irelevantní. Konkrétně se jedná o emise CO<sub>2</sub>, kvalita univerzit, znečištění PM 2.5, diskriminace menšin, předčasná úmrtí na nepřenosné nemoci, ochrana biomu, citovatelné dokumenty.

Na základě faktorového skóre lze posoudit, jak si jednotlivé země vedou v rámci dvou identifikovaných faktorů. Země jako Norsko, Švédsko, Kanada nebo Nový Zéland vykazují vysoké skóre v obou faktorech, což naznačuje nejen vysokou kvalitu života a dostupnost základních služeb (Faktor 1), ale také silné občanské svobody a





Obr. 4: Graf faktorových zátěží

### 3.2 Faktorová analýza s rotací varimax (3 faktory)

V případě faktorové analýzy s rotací varimax pro 3 faktory, se první dva faktory jeví podobné jako v předchozím případě. Co se týče třetího faktoru, tak u většiny proměnných vykazovaly zátěže hodnoty menší než  $\pm 0.5$  (viz. Tab. 1), tudíž nevysvětluje žádné relevantní proměnné.

Volba dvou faktorů se ukázala jako optimální. Dodatečný třetí faktor v alternativním řešení neposkytl výrazně lepší interpretaci ani nevysvětlil významně vyšší variabilitu. Proto bylo rozhodnuto pro dvoufaktorovou strukturu.

Tab. 1: Přehled faktorových zátěží pro 3 faktory

Faktor 1	Faktor 2	Faktor 3
0,91	-0,06	0,07
0,82	-0,23	-0,30
0,89	-0,08	-0,15
0,92	-0,12	-0,14
0,95	-0,06	0,05
0,89	-0,16	-0,24
0,88	-0,19	-0,24
-0,93	0,08	0,00
0,86	-0,28	-0,21
0,61	-0,31	-0,14
-0,42	0,77	0,15
-0,56	0,25	0,56
0,72	-0,10	-0,09
-0,70	0,29	0,26
-0,10	0,94	-0,04
-0,77	0,27	0,45
-0,66	0,17	0,08
-0,65	0,30	0,52
-0,86	0,26	0,38
0,49	-0,28	-0,34
-0,78	0,29	0,26
-0,15	-0,11	0,04
0,34	-0,45	-0,13
-0,01	0,43	0,14
-0,19	-0,56	-0,24
-0,33	0,88	0,11
-0,11	0,98	-0,02
0,02	0,77	0,08
-0,35	0,73	0,32
-0,46	0,65	0,03
0,82	-0,17	-0,30
-0,47	0,48	0,61
0,71	-0,08	-0,24
-0,17	0,63	0,34
-0,22	0,73	0,14
-0,21	0,70	0,23
0,25	-0,39	-0,44
-0,22	0,13	0,15
-0,42	0,39	0,63
-0,77	0,16	0,36

## 4 Závěr

Cílem této práce bylo analyzovat index sociálního pokroku (SPI) pomocí analýzy hlavních komponent (PCA) a následné faktorové analýzy, za účelem identifikace hlavních dimenzí, které vysvětlují rozdíly mezi státy v roce 2016. Po úvodní úpravě dat – odstranění chybějících hodnot a standardizaci – byly provedeny všechny potřebné kroky k validaci vhodnosti použití vícerozměrných metod. Detekce odlehklých hodnot a korelační analýza potvrdily opodstatnění zvoleného přístupu.

Výsledky PCA ukázaly, že dvě hlavní komponenty vysvětlují přibližně 66 % variability mezi státy. První komponenta souvisí s celkovou úrovní sociálního rozvoje, zatímco druhá rozlišuje mezi strukturálně odlišnými typy společností – zejména v oblastech svobody a participace. Vizualizace skóre komponent umožnila identifikovat typologii zemí od vysoce rozvinutých a demokratických po rozvojové země se strukturálními problémy.

Faktorová analýza s rotací varimax následně potvrdila vhodnost dvoufaktorového řešení. První faktor odráží základní lidské potřeby a kvalitu života – přístup ke zdravotní péči, vzdělání, infrastruktura a snížené míry úmrtnosti. Druhý faktor reprezentuje osobní svobody, občanská a politická práva, rovnost a možnosti participace. Identifikované faktory se vzájemně významně nekorelují, což potvrzuje, že různé aspekty sociálního pokroku se mohou v jednotlivých zemích vyvíjet nezávisle.

Pokus o trojfaktorové řešení nepřinesl výraznější zlepšení interpretace ani nárůst vysvětlené variability, a proto bylo dvoufaktorové řešení vyhodnoceno jako optimální.

## Přílohy

Indikátor	Zápis
Úmrtí na infekční nemoci (úmrtí/100 000)	Infekční úmrtnost
Zakrnění růstu dítěte (% dětí)	Zakrnění růstu dětí
Úmrtnost matek (úmrtí/100 000 živě narozených dětí)	Úmrtnost matek
Dětská úmrtnost (úmrtí/1 000 živě narozených dětí)	Dětská úmrtnost
Úmrtí související s nebezpečnou vodou, sanitací a hygienou (na 100 000 obyvatel)	Úmrtnost voda/hygiena
Populace využívající nebezpečné nebo nekvalitní zdroje vody (%)	Riziková voda
Populace používající nebezpečnou nebo nezlepšenou hygienu (%)	Riziková hygiena
Přístup k elektřině (% pop.)	Přístup k elektřině
Úmrtí související se znečištěním ovzduší v domácnostech (úmrtí/100 000)	Úmrtnost (ovzduší)
Úmrtí v dopravě (úmrtí/100 000)	Úmrtí v dopravě
Politické zabíjení a mučení (0 = nízká svoboda; 1 = vysoká svoboda)	Politické zabíjení a mučení
Přístup ke kvalitnímu vzdělání (0=nestejný; 4=stejný)	Kvalitní vzdělání
Ženy bez vzdělání	Negramotnost žen
Přístup k online správě (0=nízká; 1=vysoká)	Online správa
Cenzura médií (0=časté; 4=vzácné)	Cenzura médií
Uživatelé internetu (% pop.)	Uživatelé internetu
Předplatné mobilního telefonu (předplatné/100 osob)	Mobilní připojení
Přístup ke kvalitní zdravotní péči (0=nestejný; 4=stejný)	Zdravotní péče
Přístup k základním službám (0 = žádné; 100 = úplné pokrytí)	Základní služby
Předčasná úmrtí na nepřenosné nemoci (úmrtí/100 000)	Nepřenosná onemocnění (úmrtnost)
Očekávaná délka života 60 (let)	Délka života (v 60)
Emise skleníkových plynů (celkové ekvivalenty CO <sub>2</sub> )	Emise CO <sub>2</sub>
Populační vážené průměrné úrovně jemných částic menších než 2,5 mikronu v průměru (PM <sub>2,5</sub> )	Znečištění PM <sub>2.5</sub>
Ochrana biomu	Ochrana biomu
Úmrtí, která lze připsat znečištění venkovního ovzduší (úmrtí/100 000)	Úmrtí kvůli znečištění ovzduší
Politická práva (0=žádná práva; 40=úplná práva)	Politická práva
Svoboda projevu (0=žádná svoboda; 1=úplná svoboda)	Svoboda projevu
Svoboda vyznání (0=žádná svoboda; 4=úplná svoboda)	Svoboda vyznání
Přístup ke spravedlnosti (0=neexistující; 1=pozorováno)	Přístup ke spravedlnosti
Vlastnická práva pro ženy (0 = žádné právo; 5 = plná práva)	Vlastnická práva žen
Zranitelné zaměstnání (% zaměstnanců)	Zranitelné zaměstnání
Korupce (0=vysoké; 100=nízké)	Korupce
Předčasné manželství (% žen)	Předčasné manželství
Rovnost politické moci podle socioekonomického postavení (0=nerovná moc; 4=stejná moc)	Politická rovnost (příjem)

Rovnost politické moci podle sociálních skupin (0=nerovná moc; 4=stejná moc)	Politická rovnost (skupiny)
Rovnost politické moci podle pohlaví (0=nerovná moc; 4=stejná moc)	Politická rovnost (pohlaví)
Diskriminace a násilí vůči menšinám (0 = nízké; 10 = vysoké)	Diskriminace menšin
Univerzity vážené kvalitou (body)	Kvalita univerzit
Citovatelné dokumenty	Citovatelné dokumenty
Ženy s vyšším vzděláním (%)	Ženy s VŠ vzděláním