

UNIVERZITA PALACKÉHO V OLOMOUCI
PŘÍRODOVĚDECKÁ FAKULTA

BAKALÁŘSKÁ PRÁCE

Aplikace pro analýzu vybraných indikátorů
průmyslového rozvoje



Katedra matematické analýzy a aplikací matematiky

Vedoucí bakalářské práce: **Mgr. Kamila Fačevicová, Ph.D.**

Vypracoval(a): **Jaroslav Vymazal**

Studijní program: B1103 Aplikovaná matematika

Studijní obor Matematika–ekonomie se zaměřením na bankovníctví/pojišťovnictví

Forma studia: prezenční

Rok odevzdání: 2023

BIBLIOGRAFICKÁ IDENTIFIKACE

Autor: Jaroslav Vymazal

Název práce: Aplikace pro analýzu vybraných indikátorů průmyslového rozvoje

Typ práce: Bakalářská práce

Pracoviště: Katedra matematické analýzy a aplikací matematiky

Vedoucí práce: Mgr. Kamila Fačevicová, Ph.D.

Rok obhajoby práce: 2023

Abstrakt: Součástí softwaru R je balíček Shiny, který umožňuje jeho uživatelům vytvářet interaktivní aplikace. Cílem této práce je vytvořit s jeho pomocí aplikaci, která bude sloužit jako analytický nástroj pro UNIDO (United Nations Industrial Development Organization), pomocí kterého bude možné pohodlně sledovat vývoj ekonomických ukazatelů z hlediska času, prostoru a vzájemných vztahů.

Klíčová slova: software R, balíček Shiny, webová aplikace, bodový graf, LOESS regrese, průmyslový rozvoj

Počet stran: 48

Počet příloh: 0

Jazyk: český

BIBLIOGRAPHICAL IDENTIFICATION

Author: Jaroslav Vymazal

Title: Application for analysis of industrial development indicators

Type of thesis: Bachelor's

Department: Department of Mathematical Analysis and Applications of Mathematics

Supervisor: Mgr. Kamila Fačevicová, Ph.D.

The year of presentation: 2023

Abstract: R software includes the Shiny package, which allows its users to create interactive web applications. The goal of this thesis is to create an application using the Shiny package that will serve as an analytical tool for UNIDO (United Nations Industrial Development Organization) and will allow the user to conveniently monitor the development of economic indicators in terms of time, space and mutual relationships.

Key words: R software, R Shiny package, web application, scatter plot, LOESS regression, industrial development

Number of pages: 48

Number of appendices: 0

Language: Czech

Prohlášení

Prohlašuji, že jsem bakalářskou práci zpracoval samostatně pod vedením paní Mgr. Kamily Fačevicové, Ph.D. a všechny použité zdroje jsem uvedl v seznamu literatury.

V Olomouci dne

.....

podpis

Obsah

Úvod	7
1 Sledované ekonomické ukazatele	8
1.1 Ukazatele životní úrovně	8
1.2 Ukazatele průmyslového rozvoje	18
2 Zobrazovací a statistické metody použité v aplikaci	23
2.1 Zobrazovací metody	23
2.2 Lokální polynomiální regrese	26
3 Aplikace	30
3.1 Záložka Documentation	30
3.2 Záložka Relationship	32
3.3 Záložka Progress in Time	36
3.4 Záložka Global Overview	39
3.5 Záložka Comparison	42
Závěr	47
Literatura	48

Poděkování

Rád bych zde poděkoval Mgr. Kamile Fačevicové, PhD. za odborné vedení práce, ochotu a cenné rady, které mi při zpracovávání práce poskytla. Dále bych také chtěl poděkovat Mgr. Petře Kynčlové, PhD. za její rady poskytnuté při tvorbě aplikace.

Úvod

Balíček Shiny, který je součástí softwaru R, umožňuje jeho uživatelům vytvářet interaktivní aplikace. Cílem práce je vytvořit s pomocí tohoto balíčku aplikaci, která bude sloužit jako analytický nástroj pro UNIDO (United Nations Industrial Development Organization), pomocí kterého bude možné pohodlně sledovat vývoj ekonomických ukazatelů z hlediska času, prostoru a vzájemných vztahů. Aplikace bude navazovat na report [5] publikovaný právě organizací UNIDO, ve kterém je analyzován vztah mezi kvalitou života populace a průmyslovým rozvojem zemí ve světě.

Kromě balíčku Shiny bude při vytváření aplikace hrát velkou roli i balíček Plotly, díky kterému je možné pomocí softwaru R vytvářet interaktivní grafy. Dalším z cílů práce také je, aby byla aplikace napojena na API Světové banky a data o ukazatelech potřebná k vytváření vizualizací si prostřednictvím tohoto rozhraní brala přímo z databáze Světové banky.

Práce je v zásadě rozdělena na teoretickou část, tvořenou kapitolami 1 a 2, a praktickou část, tvořenou kapitolou 3. V teoretické části jsou rozebírány ekonomické ukazatele, které se v aplikaci objevují a statistické a vizualizační metody, kterých aplikace využívá. V části praktické je pak popis samotné aplikace a jejích funkcí.

Kapitola 1

Sledované ekonomické ukazatele

Aplikace vyvíjená v rámci této práce slouží k vizualizaci dat o různých ekonomických ukazatelech. V první kapitole budou tyto ukazatele představeny a popsány. Ke každému z ukazatelů je zde uvedena i mapa, kde jsou znázorněny hodnoty daného ukazatele ve světě v posledním roce s reprezentativními daty. Pokud v některé z map pro některou zemi figuruje bílé místo, znamená to, že v daném roce pro ni nejsou dostupná data. Údaje v závorkách u názvů ukazatelů odpovídají názvům pod nimiž se vyskytují v aplikaci. Informace obsažené v této kapitole jsou čerpány především z [1], dále pak také z [2], [3] a [4].

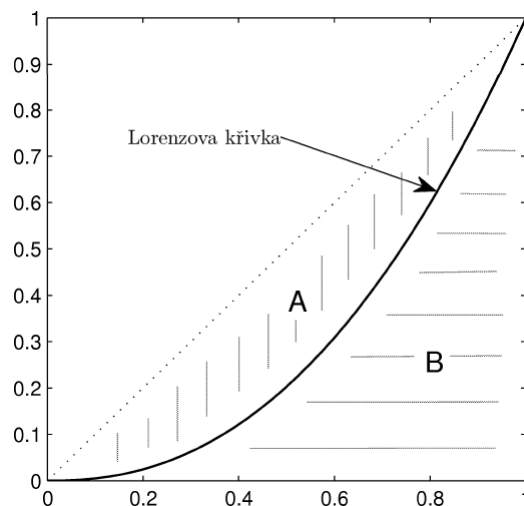
1.1. Ukazatele životní úrovně

Giniho index (Gini index)

Giniho index je procentuálním vyjádřením Giniho koeficientu, který se používá k měření příjmové nerovnosti ve společnosti. Popisuje do jaké míry se rozdělení příjmů domácností nebo jednotlivců ve společnosti odchyluje od rovnoměrného rozdělení.

Giniho koeficient můžeme graficky znázornit na jednotkovém čtverci pomocí Lorezovy křivky, což je křivka vyjadřující poměr kumulativního bohatství spodních x procent populace k celkovému bohatství. Dále osa prvního a třetího kvadrantu grafu zobrazuje rovnoměrné rozdělení příjmů ve společnosti. Giniho koeficient je pak definován jako poměr obsahu plochy ohraničené přímkou znázorňující rov-

noměrné rozdělení příjmů a Lorenzovou křivkou (A) a celkového obsahu oblasti pod zmiňovanou přímkou ($A + B$). Celou situaci názorně popisuje obrázek 1.1.

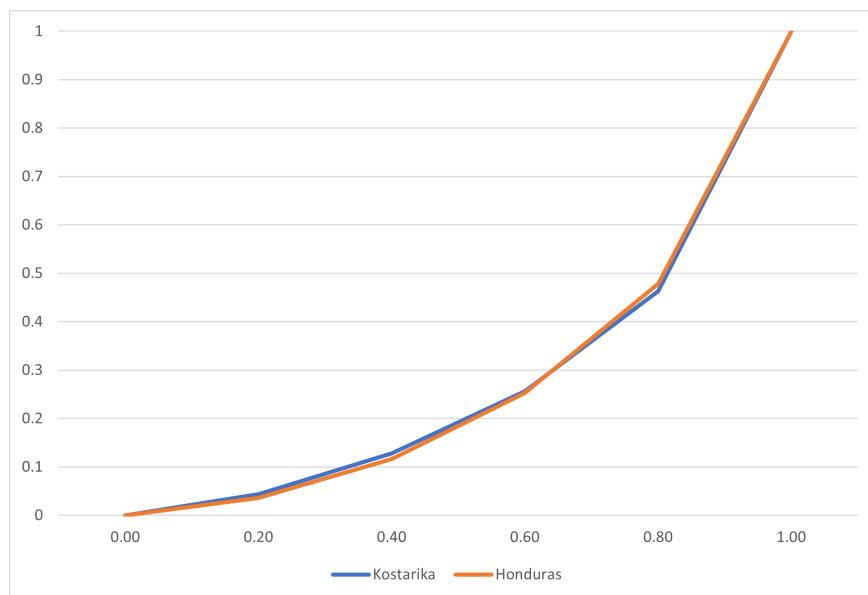


Obrázek 1.1: Grafické znázornění významu Giniho koeficientu, zdroj [2].

Podle obrázku tedy pro Giniho koeficient (G) platí:

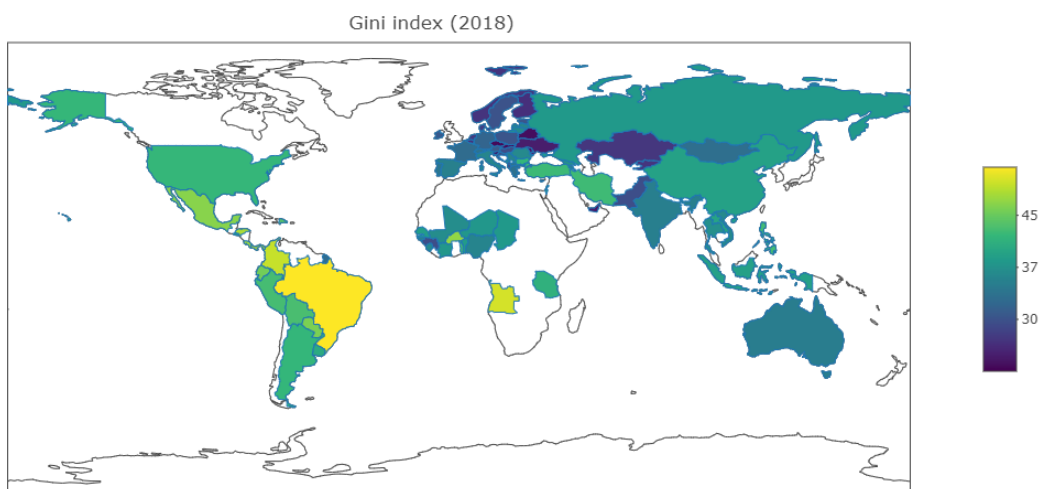
$$G = \frac{A}{A + B}.$$

Ze vzorce je patrné, že G nabývá hodnot z intervalu $\langle 0, 1 \rangle$, přičemž hodnota rovna 0 značí perfektně rovnoměrné rozdělení příjmů ve společnosti a hodnota 1 naopak naprosto nerovnoměrné rozdělení. Toto analogicky platí i pro Giniho index, tzn. hodnota 0 %, resp. 100, % značí naprosto rovnoměrné, resp. naprosto nerovnoměrné, rozdělení příjmů ve společnosti. Z obrázku také plyne jedna z nevýhod Giniho koeficientu - pro různé Lorenzovy křivky můžeme dostat stejné hodnoty tohoto ukazatele. Toto ilustruje obrázek 1.2, kde jsou znázorněny Lorenzovy křivky pro Kostariku a Honduras pro rok 2019. Obě země vykazovaly v tomto roce stejnou hodnotu Giniho indexu (48.2), ale jejich Lorenzovy křivky jsou rozdílné.



Obrázek 1.2: Porovnání Lorenzových křivek Kostariky a Hondurasu, 2019, zdroj dat [10].

Na obrázku 1.3 je vidět jakých hodnot Giniho index ve světě nabývá. Jeho hodnoty se pohybují zhruba mezi 25 a 55 procenty, přičemž nejvyšší hodnoty jsou obecně k vidění v Severní a Jižní Americe a nejnižší pak v zemích střední, východní a severní Evropy. Data jsou z roku 2018.

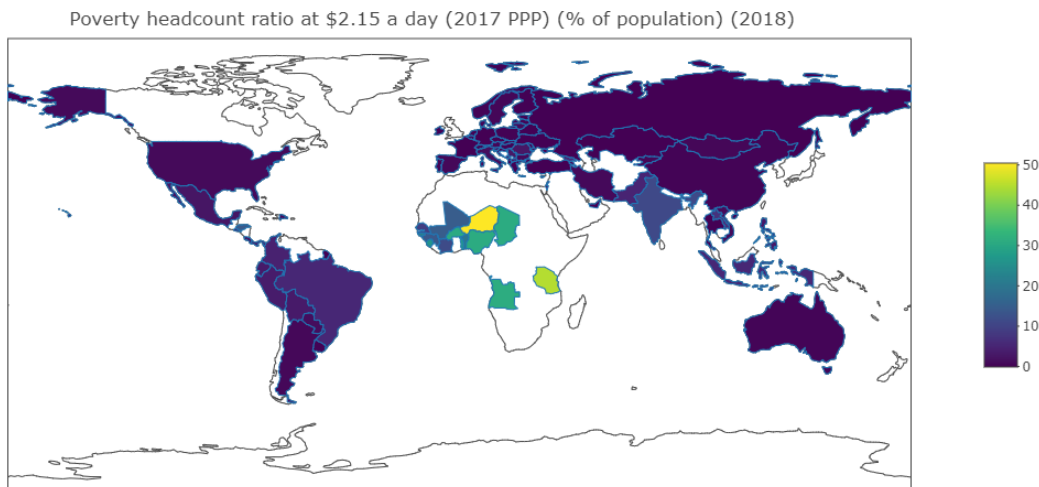


Obrázek 1.3: Hodnoty Giniho indexu ve světě, 2018

Podíl osob žijících v chudobě (Poverty headcount ratio)

Tento ukazatel vyjadřuje, jaké procento dané populace žije na denním příjmu menším než stanovená hranice chudoby. V datech Světové banky, která aplikace používá, je hranice chudoby stanovena na denním příjmu 2.15 \$, přičemž tato hodnota je poté upravena dle parity kupní síly (PPP) dané lokální měny. Tato úprava umožňuje provádět na základě tohoto ukazatele relevantní mezinárodní srovnávání. V používaných datech uvažujeme PPP z roku 2017.

Z obrázku 1.4 je patrné, že v drtivé většině zemí světa nežije v chudobě více než 10 % populace, přičemž ve vyspělých zemích hodnota tohoto ukazatele nepřekračuje 1 %. Výjimku můžeme spatřovat v Africe, kde pozorujeme hodnoty překračující 30 % a v několika zemích dokonce podíl osob žijících v chudobě nabývá hodnot okolo 50 %. Data jsou opět z roku 2018.



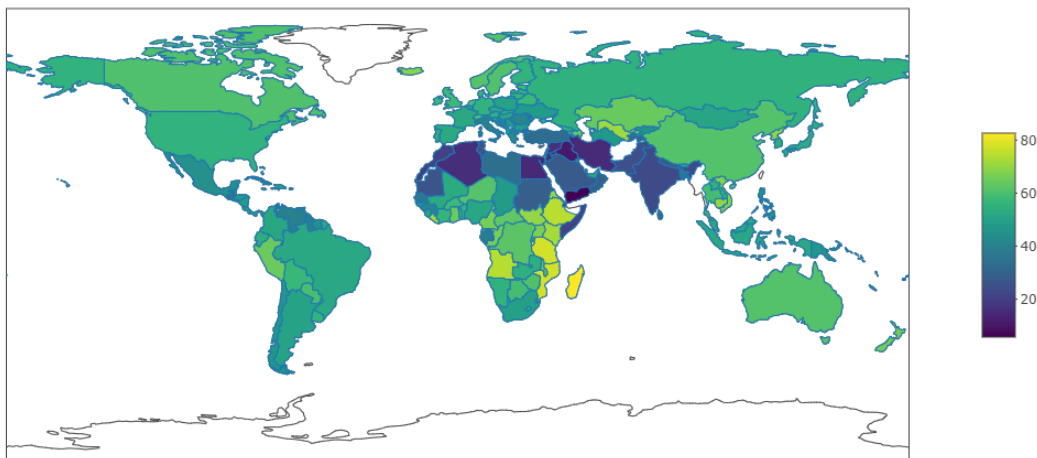
Obrázek 1.4: Podíl osob žijících v chudobě ve světě, 2018

Míra zaměstnanosti ženské pracovní síly (Female labor force participation rate)

Míra zaměstnanosti ženské pracovní síly vyjadřuje jaké procento ženské pracovní síly je ekonomicky aktivní, tzn. vykonává práci vedoucí k produkci statků a služeb. Pojmem pracovní síla se pak obecně rozumí lidé, kteří pracují, nebo si hledají práci. Neřadíme sem tedy osoby, které jsou nezaměstnané dobrovolně.

Z obrázku 1.5 je patrné, že hodnoty tohoto ukazatele ve většině zemí světa překračují 50 % a nejvyšší hodnoty můžeme pozorovat v zemích střední a jižní Afriky, kde se objevují hodnoty překračující 70 %. Nízké hodnoty (30 % a méně) pak lze spatřovat v zemích, kde je nejrozšířenějším náboženstvím islám, především tedy v zemích severní Afriky a Blízkého východu. Data jsou z roku 2021.

Labor force participation rate, female (% of female population ages 15+) (modeled ILO estimate) (2021)



Obrázek 1.5: Míra zaměstnanosti ženské pracovní síly ve světě, 2021

Podíl osob nastupujících do školy (School enrollment)

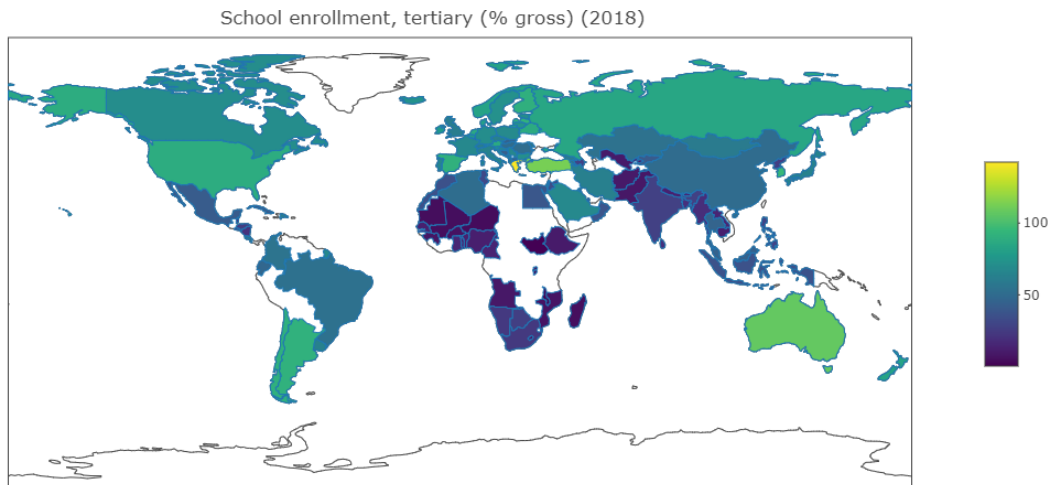
U tohoto ukazatele rozlišujeme jeho hrubou a čistou variantu. Hrubý podíl nebere ohled na věk nastupujícího. Je to podíl všech osob nastupujících v určitém roce do dané úrovně vzdělávání a populace všech lidí jejichž věk odpovídá oficiálnímu věku nástupu do dané úrovně vzdělávání. Naopak čistý podíl uvažuje místo všech nastupujících osob pouze ty z nich, jejichž věk odpovídá oficiálnímu věku nástupu do dané úrovně vzdělávání.

V databázi Světové banky se podíl vyskytuje ve třech podobách, přičemž každá odpovídá jednomu stupni vzdělávání. Rozlišujeme tedy variantu pro primární vzdělávání (v České republice zastoupeno v podobě základních škol), sekundární vzdělávání (v ČR střední školy) a terciární vzdělávání (v ČR vysoké školy).

V aplikaci se konkrétně objevuje čistý podíl osob nastupujících do primárního vzdělávání (*School enrollment, primary (% net)*), čistý podíl osob nastupujících

do sekundárního vzdělávání (*School enrollment, secondary (% net)*) a hrubý podíl osob nastupujících do terciárního vzdělávání (*School enrollment, tertiary (% gross)*).

Na obrázku 1.6 jsou znázorněny hodnoty hrubého podílu osob nastupujících do terciárního vzdělávání, jelikož ze všech tří typů podílu objevujících se v aplikaci jsou na tomto nejvíce vidět rozdíly mezi jednotlivými státy. Je vidět, že hodnoty tohoto ukazatele jsou vyšší ve vyspělejších zemích, kde zpravidla překračují 60 %. Nízké hodnoty (nižší než 25 %) můžeme pozorovat v jižní Asii a Africe (vyjma severní Afriky). V Austrálii, Turecku a Řecku jsou na obrázku patrné hodnoty vyšší než 100 %, což značí, že počet osob nastupujících do terciárního vzdělávání byl v těchto zemích v daném roce vyšší než počet osob, jejichž věk odpovídá oficiálnímu věku nástupu do terciárního vzdělávání. Data jsou z roku 2018.



Obrázek 1.6: Hrubý podíl osob nastupujících do terciárního vzdělávání ve světě, 2018

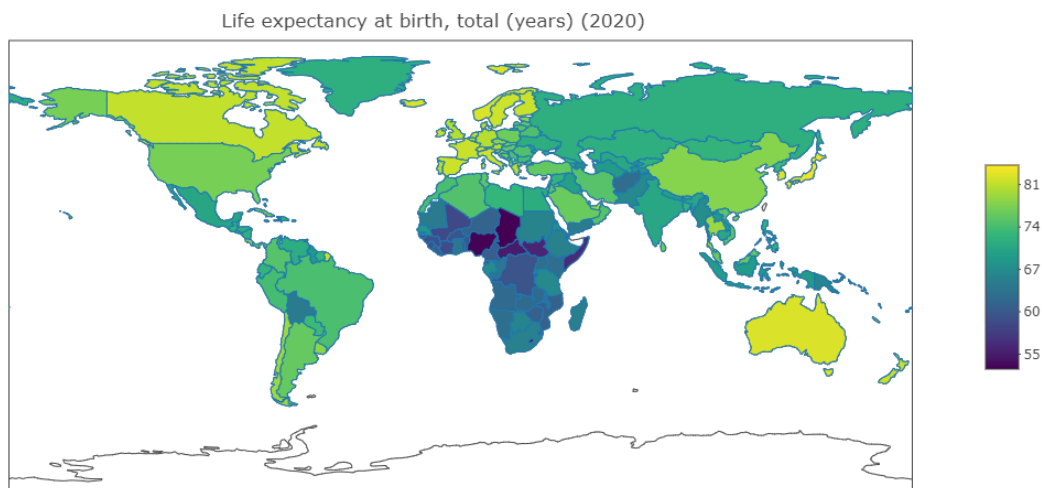
Střední délka života při narození (Life expectancy at birth)

Střední délka života při narození vyjadřuje průměrný počet let, kterého by se měl dožít novorozenec, pokud by struktura úmrtnosti v době jeho narození zůstala po celý jeho život stejná.

Tento ukazatel je zásadní pro posuzování zdraví populace v dané zemi a k

porovnávání socioekonomického rozvoje mezi zeměmi. Je třeba si však dávat pozor na fakt, že hodnotu tohoto ukazatele může výrazně snižovat vysoká úmrtnost dětí a novorozenců, která je typická pro méně rozvinuté země. V takovýchto případech pak často může docházet k tomu, že pokud jedinec přežije své dětství, dožívá se výrazně vyššího věku, než který udává hodnota očekávané délky života při narození.

Z obrázku 1.7 je patrné, že střední délka života při narození se ve většině zemí světa pohybuje nad hranicí 70 let a ve vyspělých zemích dokonce nad hranicí 75 let. Nejnižší hodnoty lze spatřovat v Africe, konkrétně např. v Čadu a Nigérii se hodnota tohoto ukazatele pohybuje těsně pod věkem 53 let. Data jsou z roku 2020.



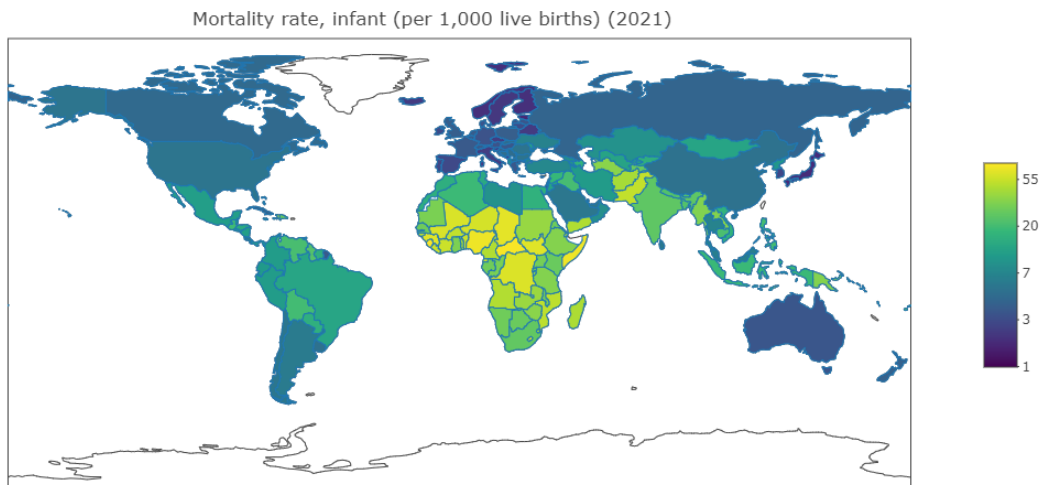
Obrázek 1.7: Střední délka života při narození ve světě, 2020

Úmrtnost novorozenců (Infant mortality rate)

Hodnota tohoto ukazatele se vykazuje ročně a je rovna relativnímu počtu dětí zemřelých před dosažením věku 1 roku, přepočítanému na tisíc živě narozených dětí.

Stejně jako střední délka života při narození, i úmrtnost novorozenců je významným ukazatelem zdraví dané populace a lze jej využít k porovnávání socioekonomického rozvoje mezi zeměmi.

Podle obrázku 1.8 nabývá tento ukazatel nejvyšších hodnot v Africe a jižní Asii. V těchto oblastech umírá před dosažením jednoho roku více než 25 z tisíce živě narozených dětí, přičemž v zemích střední Afriky pak pozorujeme úplně nejvyšší hodnoty úmrtnosti novorozenců - během prvního roku života zde v některých zemích z tisíce živě narozených dětí umírá i více než 60. Naproti tomu v ostatních zemích světa pozorujeme hodnoty nepřesahující 10 novorozeneckých úmrtí na tisíc živě narozených dětí. Když porovnáme obrázky 1.7 a 1.8, můžeme vidět, že v zemích s nižší střední délkou života při narození je vyšší úmrtnost novorozenců. Data jsou z roku 2021.



Obrázek 1.8: Úmrtnost novorozenců ve světě, 2021

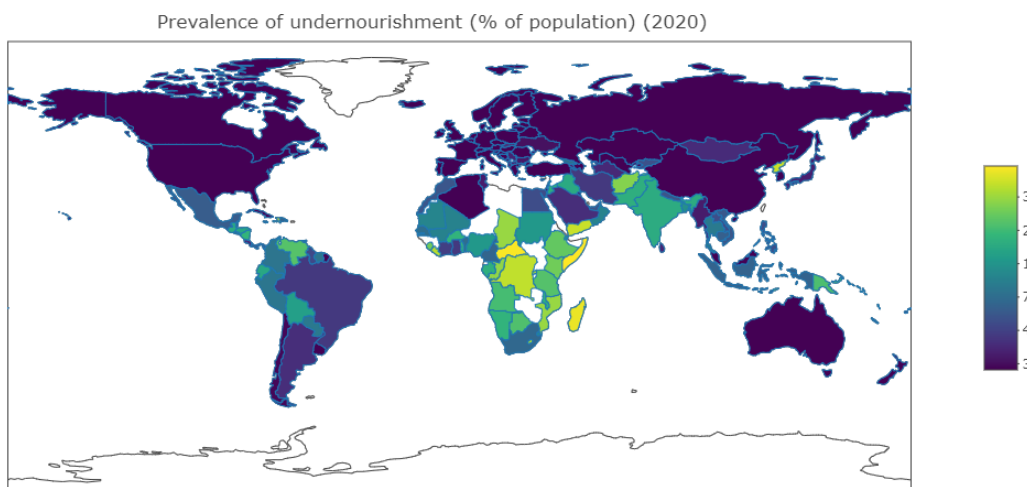
Podíl podvyživených v populaci (Prevalence of undernourishment)

Podíl podvyživených v populaci se určuje jako procento populace, jehož obvyklý příjem potravy je nedostatečný k zajištění takové energetické hladiny, která je nezbytná pro udržení normálního aktivního a zdravého života.

Výpočet hodnoty tohoto ukazatele se provádí na základě průměrného množství dostupné potravy na osobu, úrovně nerovnosti přístupu k potravě a minimálního kalorického příjmu potřebného pro průměrnou osobu.

Z obrázku 1.9 je patrné, že ve vyspělých zemích typicky trpí podvýživou méně než 5 % populace. Nejvyšší podíly podvyživených pak lze spatřovat v zemích

střední a jižní Afriky, jižní Asie a Střední a Jižní Ameriky. V těchto zemích jsou hodnoty podílu podvyživených vyšší než 10 % a v některých případech dokonce vyšší než 30 %. Ve Středoafričské republice a v Somálsku je pak situace okolo potravy úplně nejhorší a podvýživou zde trpí více než 50 % populace. Data jsou z roku 2020.



Obrázek 1.9: Podíl podvyživených v populaci napříč světovými státy, 2020

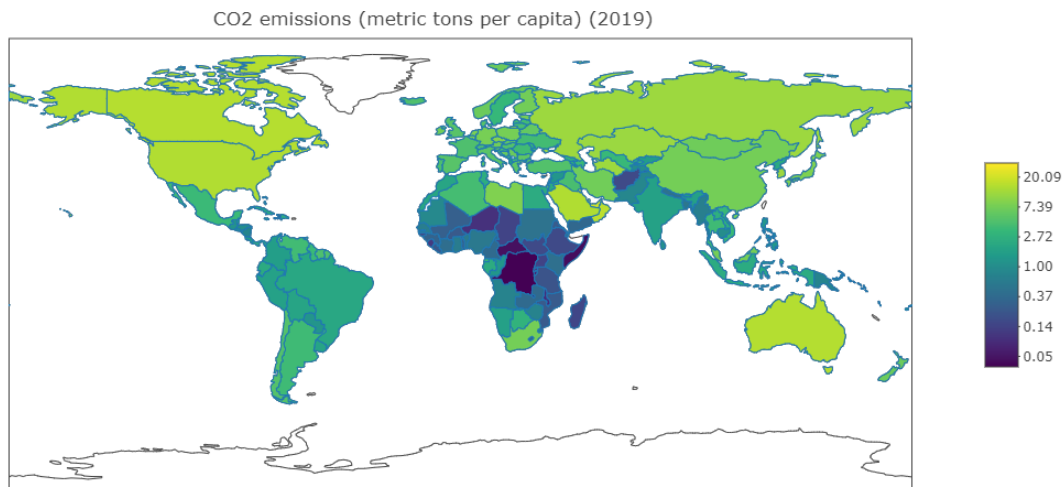
Emise oxidu uhličitého (CO_2 emissions)

Hodnota tohoto ukazatele se udává v tunách na osobu. Emise oxidu uhličitého jsou v tomto případě ty, které pocházejí ze spalování fosilních paliv a výroby cementu.

Oxid uhličitý je jedním z nejvýznamnějších skleníkových plynů, které bývají spojovány s globálním oteplováním. Z hodnoty tohoto ukazatele tedy lze získat přibližnou představu o tom, jak které země přispívají ke změnám klimatu na Zemi. Pro lepší porozumění je však vhodné brát v potaz i emise jiných skleníkových plynů, např. methanu nebo oxidu dusného.

Dle obrázku 1.10 bývá hodnota tohoto ukazatele zpravidla vyšší ve vyspělejších zemích. Nejvyšší hodnoty (přesahující 10 tun na osobu) můžeme vidět v Severní Americe, na Blízkém východě, v Austrálii, Rusku a některých zemích střední Asie. Nejnižší hodnoty (pod 3 tuny na osobu) se objevují v Africe, Střední a Jižní Ame-

rice a jižní Asii. V některých zemích střední Afriky pak můžeme dokonce vidět i hodnoty nepřesahující 100 kg na osobu. Data jsou z roku 2019.

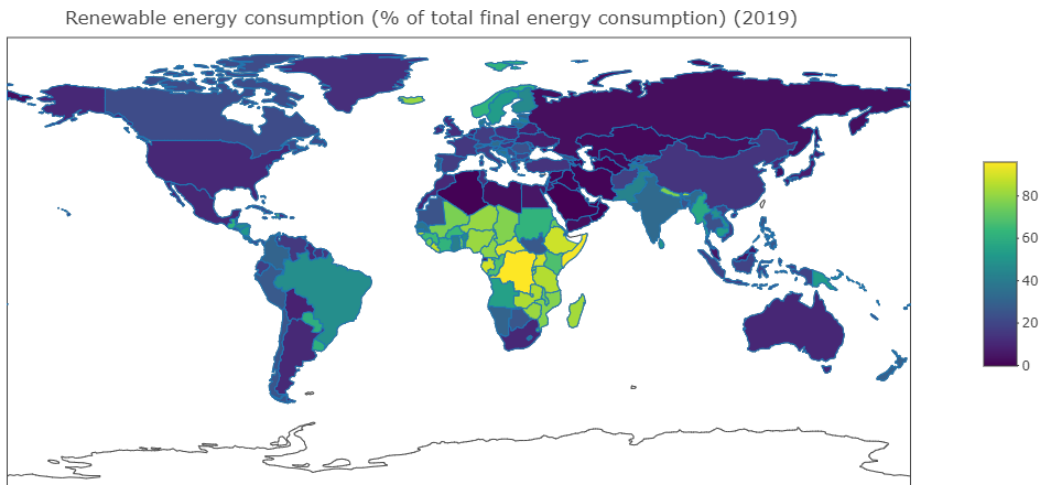


Obrázek 1.10: Hodnoty emisí oxidu uhličitého ve světě, 2019

Spotřeba energie z obnovitelných zdrojů (Renewable energy consumption)

Tento ukazatel je definován jako podíl spotřeby energie z obnovitelných zdrojů na celkové konečné spotřebě energie. Mezi obnovitelné zdroje se řadí např. voda, vítr, sluneční záření nebo biomasa.

Na obrázku 1.11 je vidět, že ve většině zemí světa zaujímá energie z obnovitelných zdrojů méně než 20 % z celkové spotřebované energie. Hodnoty vyšší než 20 % můžeme u tohoto ukazatele spatřovat v Africe, jižní Asii, některých zemích jižní Ameriky a v severní Evropě. Nejvyšší podíl na celkové spotřebě energie (více než 90 %) tvoří energie z obnovitelných zdrojů v Ugandě, Konžské demokratické republice, Somálsku a Středoafričské republice. Data jsou z roku 2019.



Obrázek 1.11: Spotřeba energie z obnovitelných zdrojů ve světě, 2019

1.2. Ukazatele průmyslového rozvoje

Hrubý domácí produkt (Gross domestic product, GDP)

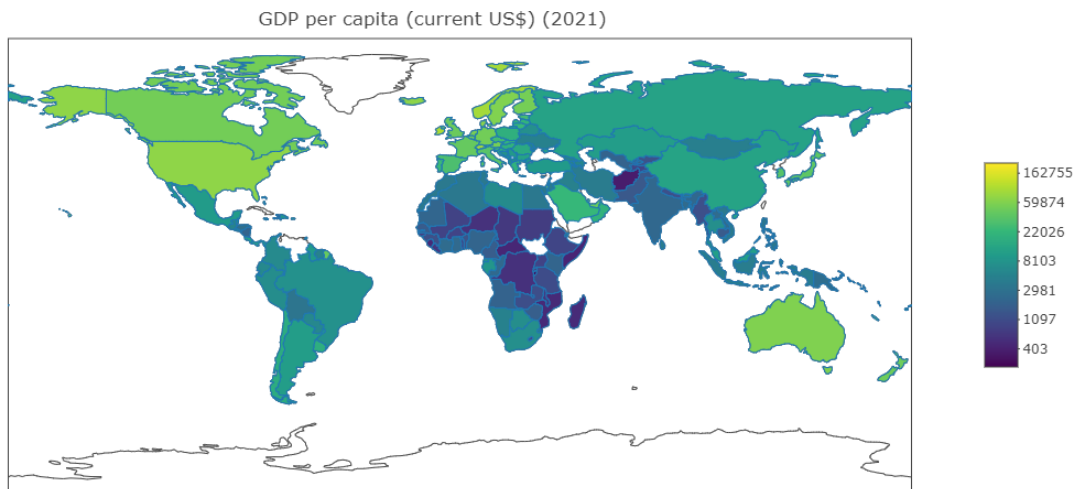
Hrubý domácí produkt (HDP) je součtem peněžních hodnot finálních (konečných) výrobků a služeb, vyprodukovaných během jednoho roku výrobními faktory alokovanými (umístěnými) v dané zemi, bez ohledu na to, kdo tyto faktory vlastní.

Pomocí hrubého domácího produktu můžeme měřit ekonomický výkon dané země. Jde o jeden z makroekonomických agregátů, což jsou souhrnné národohospodářské veličiny, pomocí nichž se měří rozsah hospodářské činnosti jednotlivých zemí.

Rozlišujeme nominální a reálný HDP. Nominální HDP je vypočten v běžných cenách, tzn. v cenách, které převládají na trhu v době, za kterou je HDP počítán. Reálný HDP je naproti tomu vypočten ve stálých cenách, tzn. v cenách očištěných od změn. Stálými cenami rozumíme ceny toho roku, který stanovíme jako výchozí.

V aplikaci se objevuje HDP na osobu, což je celkový HDP vydělený středním počtem obyvatel dané země v daném roce. Jde o nominální HDP a jeho hodnoty jsou uváděny v amerických dolarech.

Dle obrázku 1.12 nabývá HDP na osobu nejnižších hodnot (pod 5000 amerických dolarů na osobu) v zemích Afriky a střední a jižní Asie. V rozvinutých zemích se hodnota pohybuje nad hranicí 20 tisíc amerických dolarů na osobu a v Austrálii, USA, Kanadě a některých evropských zemích je HDP na osobu dokonce vyšší než 50 tisíc dolarů na osobu. Data jsou z roku 2021.



Obrázek 1.12: Hodnoty nominálního HDP na osobu ve světě, 2021

Přidaná hodnota výrobního sektoru (Manufacturing value added, MVA)

Přidaná hodnota výrobního sektoru vyjadřuje hodnotu čisté produkce všech rezidentských výrobních jednotek. Tuto hodnotu lze získat sečtením výkonů a odečtením mezispotřeby daných výrobních jednotek.

Měření MVA vyžaduje vhodné vymezení typu ekonomické činnosti a území, na kterém se činnost uskutečňuje. Hranice výroby jako ekonomické činnosti jsou definovány v klasifikaci ISIC (International Standard Industrial Classification of All Economic Activities, viz [8]).

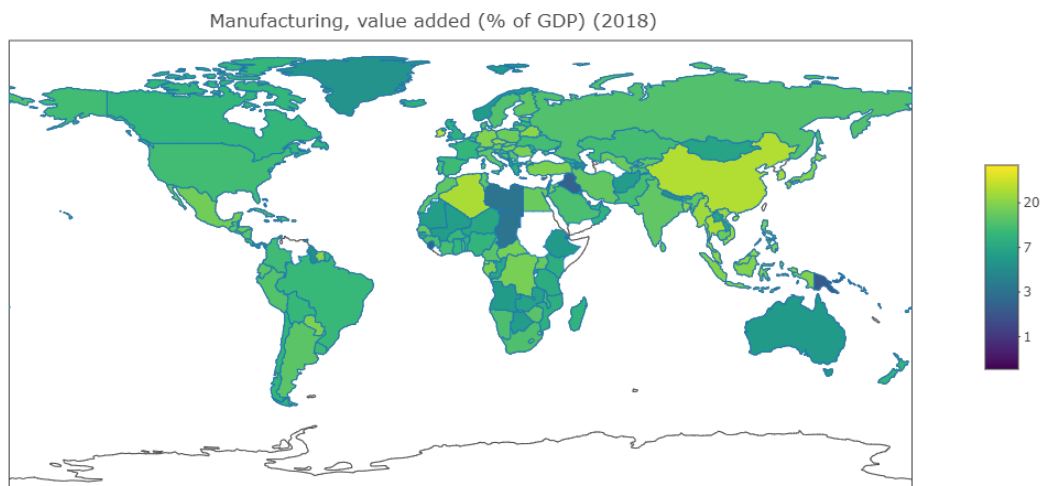
Přestože HDP poskytuje důležitou informaci o celkovém ekonomickém rozvoji země, neodhaluje žádné informace o odvětvovém složení ekonomiky ani o různých stupních průmyslového rozvoje. Právě k zachycení různých úrovní průmyslového rozvoje dané země a k měření míry tohoto rozvoje se používá ukazatel MVA, resp. MVA na osobu.

V aplikaci se přidaná hodnota výrobního sektoru objevuje ve dvou variantách:

- hodnota MVA dané země v amerických dolarech, při stálé cenové hladině z roku 2015 (*Manufacturing value added (constant 2015 US\$)*),
- hodnota MVA jako procento HDP (*Manufacturing value added (% of GDP)*).

Dále se v aplikaci objevuje také přidaná hodnota středního a high-tech výrobního sektoru (vymezeno v ISIC) jako procento celkové MVA (*Medium and high-tech manufacturing value added*).

Na obrázku 1.13 jsou znázorněny hodnoty MVA jako procenta HDP ve světě. Nejnižší hodnoty jsou k vidění v Austrálii a v některých zemích Afriky a Blízkého východu. V těchto zemích MVA tvoří méně než 5 % celkového HDP. Naproti tomu v Číně, jižní Asii a v některých afrických a evropských zemích tvoří MVA přes 20 % HDP. Data jsou z roku 2018.

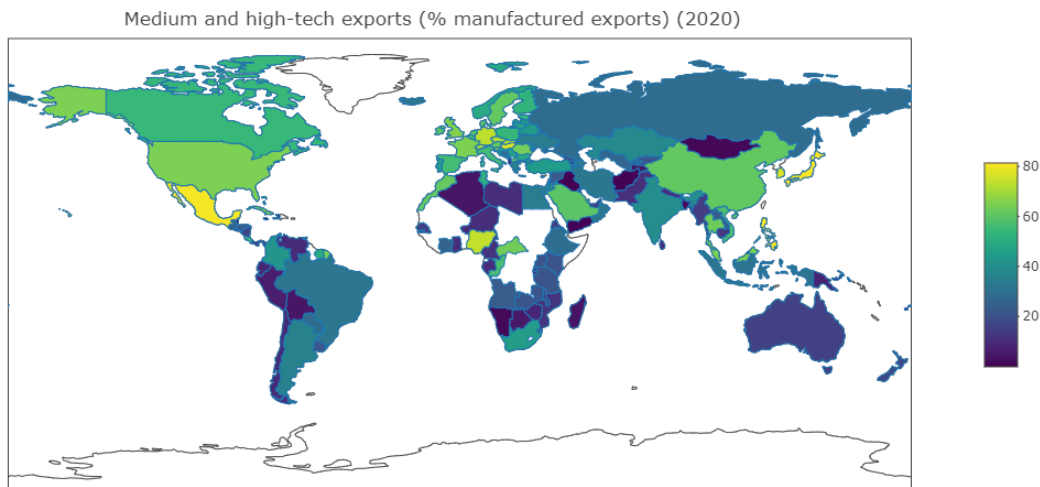


Obrázek 1.13: MVA jako procento HDP ve světě, 2018

Vývoz středních a špičkových technologií (Medium and high-tech exports)

Tento ukazatel je udáván v procentech z celkového vývozu vyrobeného zboží. Klasifikace technologií na střední, špičkové aj. je definována v klasifikaci SITC (Standard International Trade Classification, viz [9]).

Dle obrázku 1.14 jsou hodnoty tohoto ukazatele vyšší (přes 50 %) v rozvinutých zemích, typicky v Evropě, Severní Americe, Číně, Japonsku a Jižní Korei. Překvapivě nízký objem vývozu středních a špičkových technologií má Austrálie (jen lehce přes 15 %) a nejnižší hodnoty tohoto ukazatele jsou v některých zemích Jižní Ameriky, Afriky a Asie (pod 10 %). Data jsou z roku 2020.

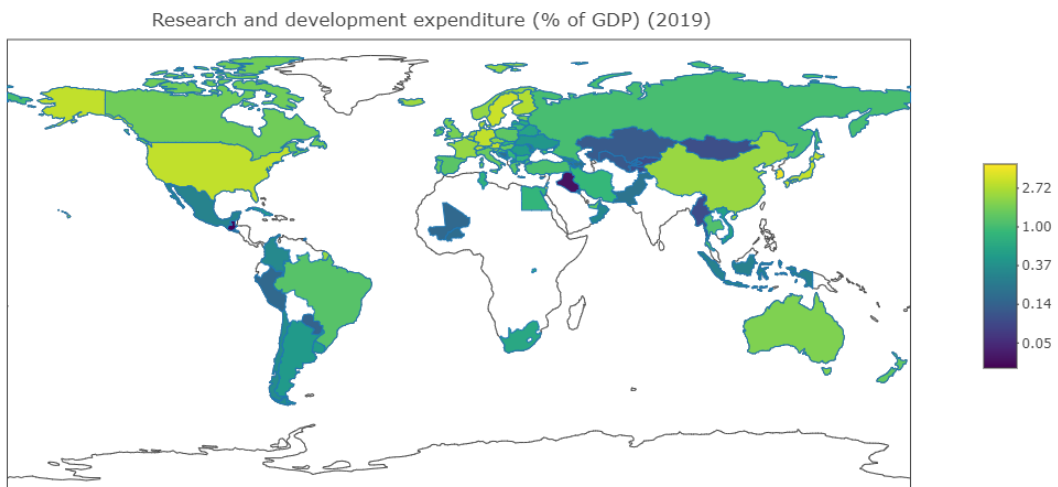


Obrázek 1.14: Hodnoty vývozu středních a špičkových technologií ve světě, 2020

Výdaje na výzkum a vývoj (Research and development expenditure)

Výdaje na výzkum a vývoj jsou v používaných datech Světové banky vyjádřeny jako procento HDP. Zahrnují kapitálové i běžné výdaje v soukromém sektoru, státním sektoru, vysokém školství a soukromých neziskových organizacích. Jsou sem zahrnovány i výdaje domácích firem na výzkum v zahraničí. Výzkum a vývoj zahrnuje teoretický i aplikovaný výzkum a experimentální vývoj.

Hodnoty tohoto ukazatele jsou klíčovým indikátorem toho, do jaké míry státní a soukromý sektor dané země usiluje o získání konkurenční výhody ve vědě a technologiích. Z obrázku 1.15 je patrné, že nejvyšší hodnoty jsou k vidění v USA, Kanadě, Evropě, Číně, Austrálii, Japonsku, Brazílii a Jižní Korei. V těchto zemích tvoří výdaje na výzkum a vývoj více než 1 % HDP, někde dokonce i více než 3 %. Data jsou z roku 2019.



Obrázek 1.15: Výdaje na výzkum a vývoj ve světě, 2019

Kapitola 2

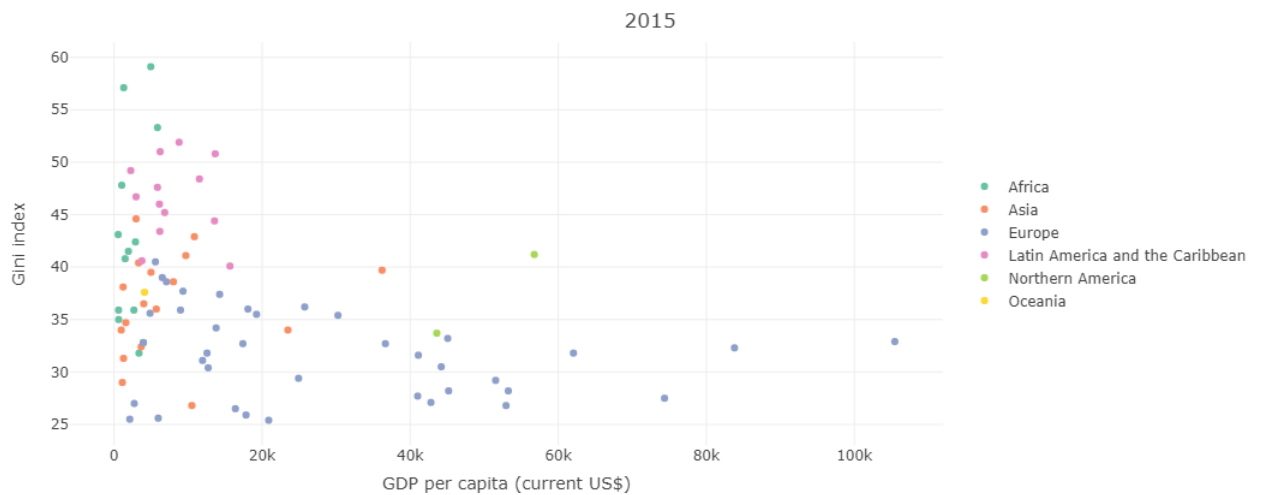
Zobrazovací a statistické metody použité v aplikaci

2.1. Zobrazovací metody

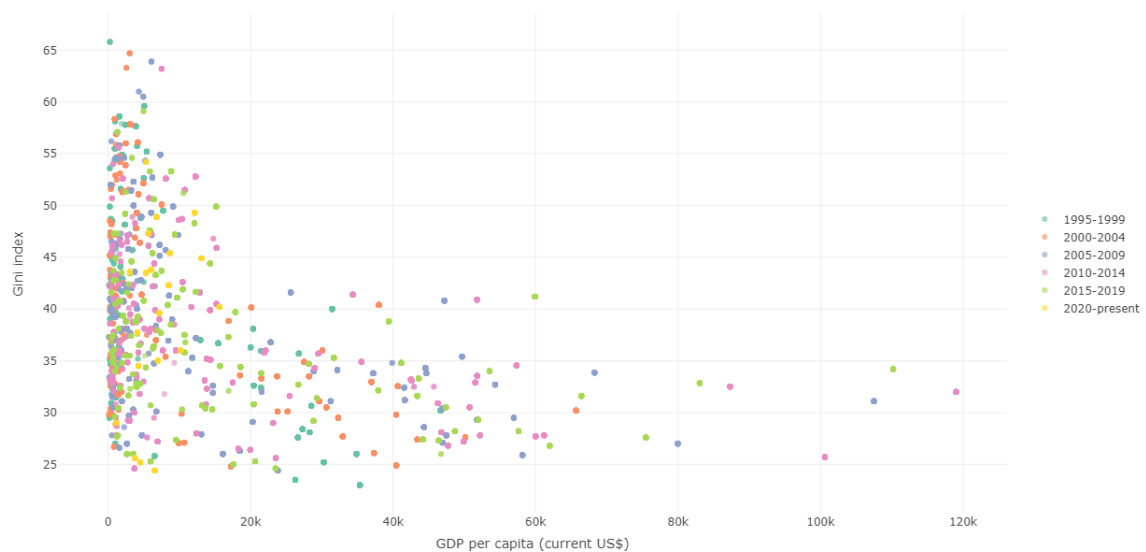
Uživatel aplikace si bude moci s její pomocí nechat vykreslit několik různých typů grafů. Ty jsou popsány v této části práce. Veškeré obrázky, které se zde objevují, jsou přímo z aplikace.

Prvním typem grafu, který aplikace umožňuje vykreslit, je *bodový graf* (angl. *scatter plot*), přičemž se zde objevují hned dvě varianty tohoto grafu. První z nich je obyčejný bodový graf sloužící k popisu vztahu dvou ukazatelů ve zvoleném roce (viz obrázek 2.1). Na každé z os jsou vynášeny hodnoty jednoho ze zvolených ukazatelů a jeden bod grafu odpovídá dvojici hodnot ukazatelů pro jeden ze světových států. Státy navíc mohou být barevně odlišeny dle příjmů jejich ekonomiky nebo dle kontinentů, na kterých leží. Druhou variantou je bodový graf, který umožňuje uživateli zkoumat vývoj vztahu dvou veličin v průběhu let (viz obrázek 2.2). Na osách grafu se v tomto případě opět objevují dvě uživatelem zvolené proměnné, ale jako body grafu vystupují dvojice hodnot příslušné dané zemi pro určitý časový úsek. Každá země se tedy v grafu vyskytuje několikrát a barevné rozlišení tentokrát odpovídá jednotlivým časovým úsekům. Hodnoty z dvojice jsou získány jako medián hodnot zvoleného ukazatele pro daný časový úsek. V aplikaci je u obou grafů také možnost transformovat hodnoty na jedné z os do logaritmického měřítka. Místo hodnot veličiny x , resp. y , se pak tedy

na danou osu vynáší hodnoty transformované veličiny $\log(x)$, resp. $\log(y)$. To je výhodné v případě, že rozdělení sledované veličiny je asymetrické s převážně nízkými hodnotami, ale nemalým počtem hodnot velkých.

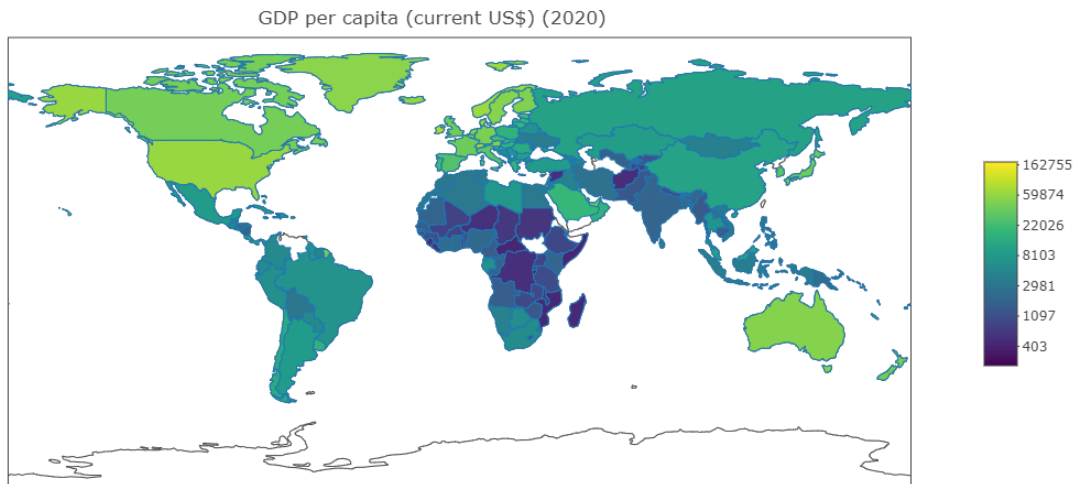


Obrázek 2.1: Příklad klasického bodového grafu



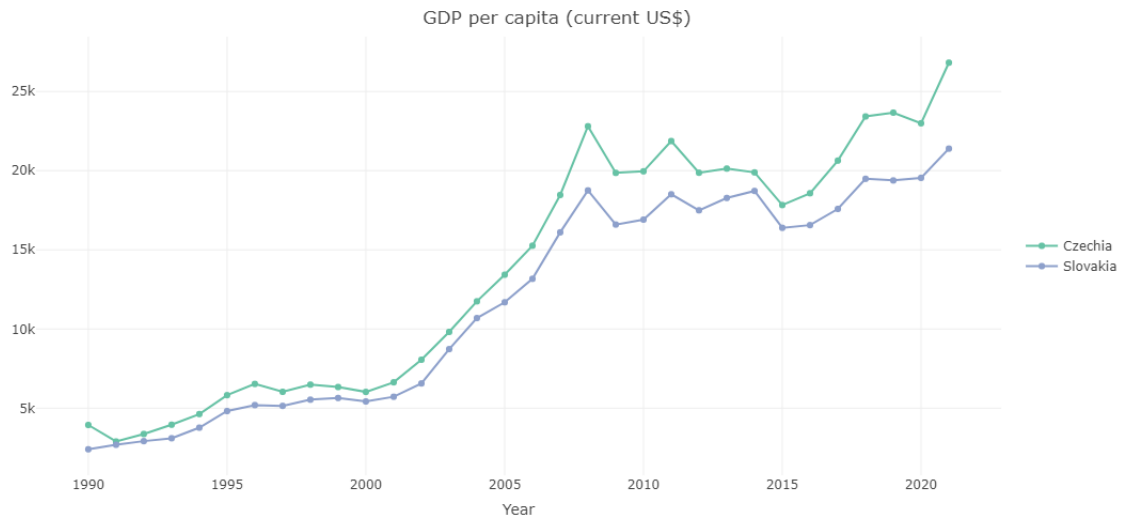
Obrázek 2.2: Příklad bodového grafu umožňující sledování vývoje vztahu dvou veličin v čase

Druhým grafem objevujícím se v aplikaci jsou *mapy*. Tento typ grafu umožňuje uživateli sledovat hodnoty zvoleného ukazatele ve zvoleném roce pro všechny státy (viz obrázek 2.3). Opět je zde také možnost transformovat hodnoty vizualizované proměnné do logaritmického měřítka.



Obrázek 2.3: Příklad mapy s logaritmickým měřítkem

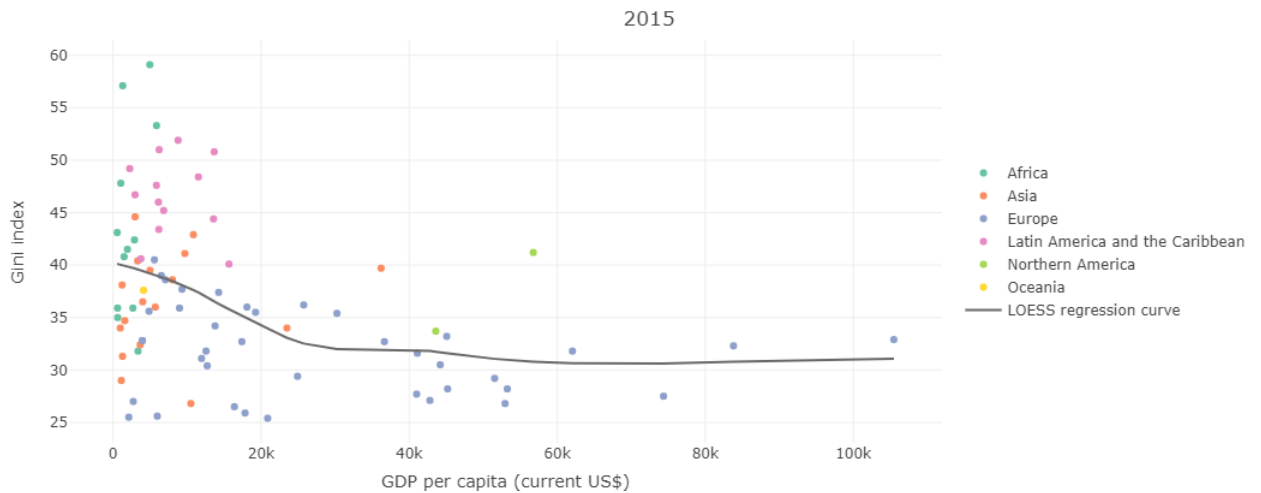
Posledním typem grafu, který je v aplikaci k vidění, je *spojnicový graf*, který uživateli umožňuje sledovat hodnoty zvoleného ukazatele pro zvolené země v průběhu let (viz obrázek 2.4). Opět lze hodnoty ukazatele transformovat do logaritmického měřítka.



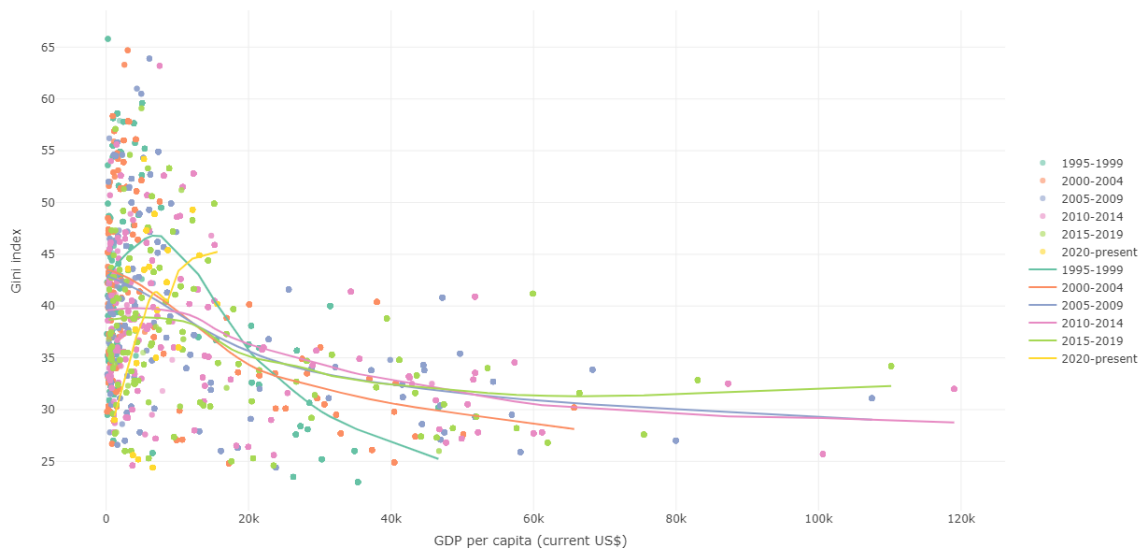
Obrázek 2.4: Příklad spojnicového grafu

2.2. Lokální polynomiální regrese

Významnou statistickou metodou objevující se v aplikaci, je lokální polynomiální regrese. V aplikaci se konkrétně používá metoda *LOESS* (*locally estimated scatterplot smoothing*). Jedná se o neparametrickou metodu regrese, která je v podstatě zobecněním metody klouzavých průměrů. V aplikaci je metoda aplikována na oba typy dříve představených bodových grafů (viz obrázky 2.5 a 2.6), přičemž uživatel si může zvolit, zda v grafu regresní křivka má být zobrazena nebo ne. Informace o konstrukci LOESS křivky byly čerpány z [6] a [7].



Obrázek 2.5: Příklad bodového grafu s regresní křivkou



Obrázek 2.6: Příklad bodového grafu s regresními křivkami

Mějme datový soubor o rozsahu n tvořený body x_i , $i = 1, \dots, n$ a k nim příslušnými hodnotami y_i . Metoda funguje tak, že ke každé hodnotě y_i datového souboru určíme vyrovnanou hodnotu \hat{y}_i pomocí metody vážených nejmenších čtverců, kterou aplikujeme pouze na určitou podmnožinu bodů datového souboru, které jsou nejbližšími sousedy bodu x_i . Označme tuto podmnožinu $N_k(x_i)$, kde k je daný počet nejbližších sousedů bodu x_i . Do této množiny řadíme i bod x_i .

Pro každý bod datového souboru x_i tedy řešíme problém:

$$\min_{\beta_0, \beta_1, \dots, \beta_m} \sum_{j: x_j \in N_k(x_i)} w(x_j) [y_j - P_m(x_j)],$$

kde $w(x_j)$ značí váhu bodu x_j , $P_m(x_j)$ hodnotu polynomu m -tého stupně v bodě x_j a $\beta_0, \beta_1, \dots, \beta_m$ jsou parametry tohoto polynomu, které pro dané x_i chceme odhadnout. Jak hodnoty vah, tak i stupeň polynomu použitého k aproximaci je třeba zvolit. Společně s nimi je pak ještě třeba určit hodnotu k , tj. počet nejbližších sousedů bodu x_i , který budeme při výpočtu uvažovat. Pro vyrovnanou hodnotu \hat{y}_i pak platí

$$\hat{y}_i = P_m(x_i).$$

Lokálně můžeme prokládat různý stupeň polynomu, přičemž běžně se používají polynomy prvního a druhého stupně, tzn. lokálně lineární a lokálně kvadratický trend. Případným použitím polynomu nultého stupně, tedy konstantní funkce, se z metody LOESS stává metoda vážených klouzavých průměrů. V aplikaci je používán lokálně lineární trend, tzn.

$$P_m(x_j) = \beta_0 + \beta_1 x_j,$$

kde β_0 a β_1 jsou parametry, které je pro dané x_i třeba odhadnout.

Velikost množiny bodů datového souboru $N_k(x_i)$, na které probíhá vyhlazení určujeme pomocí parametru λ , kterému se říká rozpětí nebo také jednoduše vyhlazovací parametr. Vyjadřuje podíl bodů datového souboru, který bude použit k lokálnímu vyhlazování. Pokud tedy např. $\lambda = 0.8$, znamená to, že bude pro výpočet vyrovnaných hodnot použito vždy 80 % bodů datového souboru, které jsou nejbližšími sousedy bodu x_i . Jelikož polynom m -tého stupně vyžaduje k odhadu parametrů $m + 1$ bodů, musí rozpětí splňovat podmínku $\lambda \in \langle \frac{m+1}{n}, 1 \rangle$, kde n je rozsah datového souboru. Platí, že čím vyšší hodnotu λ zvolíme, tím většího vyhlazení docílíme. V aplikaci je hodnota rozpětí rovna 0.8.

Váhy volíme tak, aby platilo, že čím vzdálenější je bod od toho bodu, pro který určujeme vyrovnanou hodnotu, tím menší je jeho váha. Volba váhové funkce

$w(x_j)$ je flexibilní, často se používá např. *trikubická váhová funkce*

$$w(x_j) = (1 - |d_j|^3)^3$$

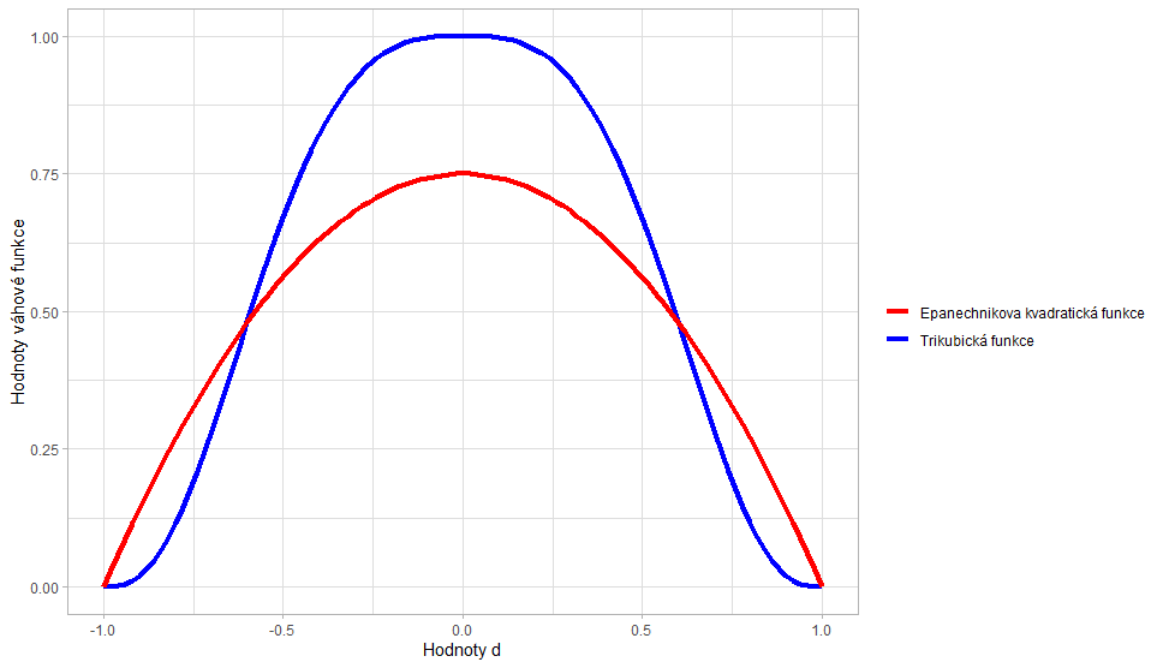
nebo *Epanechnikova kvadratická funkce*

$$w(x_j) = \frac{3}{4}(1 - d_j^2),$$

kde pro oba případy platí, že d_j je vzdálenost bodu x_j od bodu x_i , ke kterému určujeme vyrovnanou hodnotu, znormovaná tak, aby platilo $d_j \in \langle -1, 1 \rangle$. Tohoto znormování lze docílit např. následujícím způsobem:

$$d_j = \frac{x_i - x_j}{\max_{x_j \in N_k(x_i)} |x_i - x_j|}.$$

Z grafu těchto funkcí, které jsou vidět na obrázku 2.7, je patrné, že čím větší je $|d_j|$, tím menší je hodnota $w(x_j)$, tj. váha bodu x_j . V aplikaci se k určení vah používá trikubická funkce.



Obrázek 2.7: Porovnání grafů váhových funkcí

Kapitola 3

Aplikace

Aplikace byla vytvořena v jazyce R, přičemž hlavní roli při její tvorbě hrály balíčky Shiny a Plotly, a slouží k interaktivní vizualizaci dat. Veškerá data používaná k vytváření vizualizací, jsou z databáze Světové banky, na kterou je aplikace napojena prostřednictvím API.

Aplikace má celkem pět záložek, z nichž jedna je čistě informační a zbylé čtyři slouží k vizualizaci dat o ekonomických ukazatelech popsanych v první kapitole. Na každé z nich si pak uživatel může zobrazit nějaký z grafů popsanych v druhé kapitole (část 2.1.). V této kapitole budou popsány funkce aplikace a její vzhled, aplikaci samotnou pak je možné najít na https://vymaja.shinyapps.io/UNIDO_visualization_tool.

3.1. Záložka Documentation

Tato záložka funguje jako dokumentace k aplikaci. Jsou zde informace o samotné aplikaci a stručný popis sledovaných ekonomických ukazatelů a vizualizačních a statistických metod, kterých aplikace využívá. Je zde také uveden zdroj dat. Celá záložka je vidět na obrázku 3.1.

ABOUT THE APPLICATION

This application provides a user-friendly environment to visualize statistical indicators, tracking progress and performance of economies and the linkages between industrial development and the well-being of population.

In 2015, the importance of industrial development was recognized under the Sustainable Development Goal (SDG) 9 by the 2030 Agenda for Sustainable Development formulated by the United Nations General Assembly (UNGA). The Sustainable Development Goals are a universal call to action to end poverty, protect the planet and improve the lives and prospects of everyone, everywhere. The 17 Goals were adopted by all UN Member States in 2015, as part of the 2030 Agenda for Sustainable Development which set out a 15-year plan to achieve the Goals.

SDG 9 calls to build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation. Although industrialization contributes to economic growth, its impact differs depending on the country's stage of development. Industrial growth in higher-income economies will largely depend on the level of adoption of new technologies, intelligent production processes and on limiting the impact of industrial production on the local natural environment and the global climate.

Structural transformation, one of the most promising strategies for reaching sustainable economic growth and improving the population's living standards involves a stable transition of economic activity from the primary sector into manufacturing and higher value-added services, and from lower- to higher-productivity activities. In addition to its benefits to diversification and structural change, a transition to higher-technology and innovation-focused industries is essential to support green growth as these activities are generally less energy- and emission-intensive.

It is indisputable that the achievement of SDG 9 is linked to meeting the other Goals and targets of the 2030 Agenda. Inclusive and sustainable industrialization drives sustained economic growth, the creation of decent jobs and income (SDG 8); it helps reduce poverty (SDG 1), hunger (SDG 2) and inequalities (SDG 5 and 10), while improving health and well-being (SDG 3), increasing resource and energy efficiency (SDG 6, 7, 11, 12) and reducing greenhouse gas and other polluting emissions, including from chemicals (SDG 13, 14, 15).

There is strong evidence that citizens living in developed industrialized countries enjoy far more prosperous and healthy lives than those who reside in least developed countries (LDCs). The former benefit from high levels of education, better social security and health services, sophisticated transport and communication networks, and access to information, knowledge, technology and financial facilities required by businesses.

The application helps to investigate how closely industrial development is linked to people's living conditions and the quality of their lives. It provides empirical evidence on the association between industrial development and other dimensions of sustainable development, with a view to improving the understanding of these relations among policymakers at both national and international level.

For more information about how industrial development affects the well-being of a population, see [this report](#).

All plots in the application are interactive providing the user with a download option via the action button as .html, or directly via an icon in the upper right side of each plot as .png.

METHODOLOGY

The visualizations are based on three types of plots. Maps allow the user to perform a worldwide comparison. Another type of plot used is a time series format, where the user can see the progression of an indicator for selected countries in a given year. Relationships between two indicators are visualized on scatter plots. There are two types of scatter plots - the classic simple scatter plot where the user can see the relationship between two indicators in one year and a scatter plot which shows the progression of the relationship of two indicators in time. In the latter, each country is represented in the given time period by the median value reported for the relevant years. The overall trend of the relationship is indicated by the Loess smoothing curve. The Loess regression is a non-parametric technique that uses a local weighted regression to fit a smooth curve through points in a scatter plot. Loess curves can reveal trends and cycles in data that might be difficult to model with a parametric curve. In this case, the Loess regression takes 80% of the available data and locally fits a first degree polynomial for each datapoint. In extreme cases, when there is not enough data to compute the curve correctly, the curve may approach infinity or negative infinity.

THE DATA

All the data presented in the application are provided by the World Bank API. All their descriptions are sourced from the World Bank as well. The indicators included in the application are:

- *GDP per capita (current US\$)* - GDP per capita is gross domestic product divided by midyear population. GDP is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources. Data are in current U.S. dollars.
- *Manufacturing, value added (constant 2015 US\$)* - Manufacturing value added is the net output of the manufacturing sector after adding up all outputs and subtracting intermediate inputs. It is calculated without making deductions for depreciation of fabricated assets or depletion and degradation of natural resources. Data are in constant 2015 prices, expressed in U.S. dollars.
- *Manufacturing, value added (% of GDP)* - The proportion of manufacturing value added in country's GDP
- *Medium and high-tech manufacturing value added (% manufacturing value added)* - The proportion of medium and high-tech industry value added in total value added of manufacturing.
- *Medium and high-tech exports (% manufactured exports)* - Share of medium and high-tech manufactured exports in total manufactured exports.
- *Research and development expenditure (% of GDP)* - Gross domestic expenditures on research and development (R&D), expressed as a percent of GDP. They include both capital and current expenditures in the four main sectors: Business enterprise, Government, Higher education and Private non-profit. R&D covers basic research, applied research, and experimental development.
- *Gini index* - Gini index measures the extent to which the distribution of income among individuals or households within an economy deviates from a perfectly equal distribution. Gini index of 0 represents perfect equality, while an index of 100 implies perfect inequality.
- *Poverty headcount ratio at \$2.15 a day (2017 PPP) (% of population)* - Poverty headcount ratio at \$2.15 a day is the percentage of the population living on less than \$2.15 a day at 2017 purchasing power adjusted prices.
- *Labor force participation rate, female (% of female population ages 15+)* (modeled ILO estimate) - This indicator measures the proportion of the female population ages 15 and older that is economically active: all who supply labor for the production of goods and services during a specified period.
- *School enrollment, primary (% net)* - Net enrollment rate is the ratio of children of official school age who are enrolled in school to the population of the corresponding official school age. Primary education provides children with basic reading, writing, and mathematics skills along with an elementary understanding of such subjects as history, geography, natural science, social science, art, and music.
- *School enrollment, secondary (% net)* - Net enrollment rate for secondary education. Secondary education completes the provision of basic education that began at the primary level, and aims at laying the foundations for lifelong learning and human development, by offering more subject- or skill-oriented instruction using more specialized teachers.
- *School enrollment, tertiary (% gross)* - Gross enrollment ratio is the ratio of total enrollment, regardless of age, to the population of the age group that officially corresponds to the level of education shown. Tertiary education, whether or not to an advanced research qualification, normally requires, as a minimum condition of admission, the successful completion of education at the secondary level.
- *Life expectancy at birth, total (years)* - Life expectancy at birth indicates the number of years a newborn infant would live if prevailing patterns of mortality at the time of its birth were to stay the same throughout its life.
- *Mortality rate, infant (per 1,000 live births)* - Infant mortality rate is the number of infants dying before reaching one year of age, per 1,000 live births in a given year.
- *Prevalence of undernourishment (% of population)* - Prevalence of undernourishment is the percentage of the population whose habitual food consumption is insufficient to provide the dietary energy levels that are required to maintain a normal active and healthy life. Data showing as 2.5 may signify a prevalence of undernourishment below 2.5%.
- *CO2 emissions (metric tons per capita)* - Carbon dioxide emissions are those stemming from the burning of fossil fuels and the manufacture of cement. They include carbon dioxide produced during consumption of solid, liquid, and gas fuels and gas flaring.
- *Renewable energy consumption (% of total final energy consumption)* - Renewable energy consumption is the share of renewables energy in total final energy consumption.

You can find more information about the indicators and their more detailed descriptions [here](#).

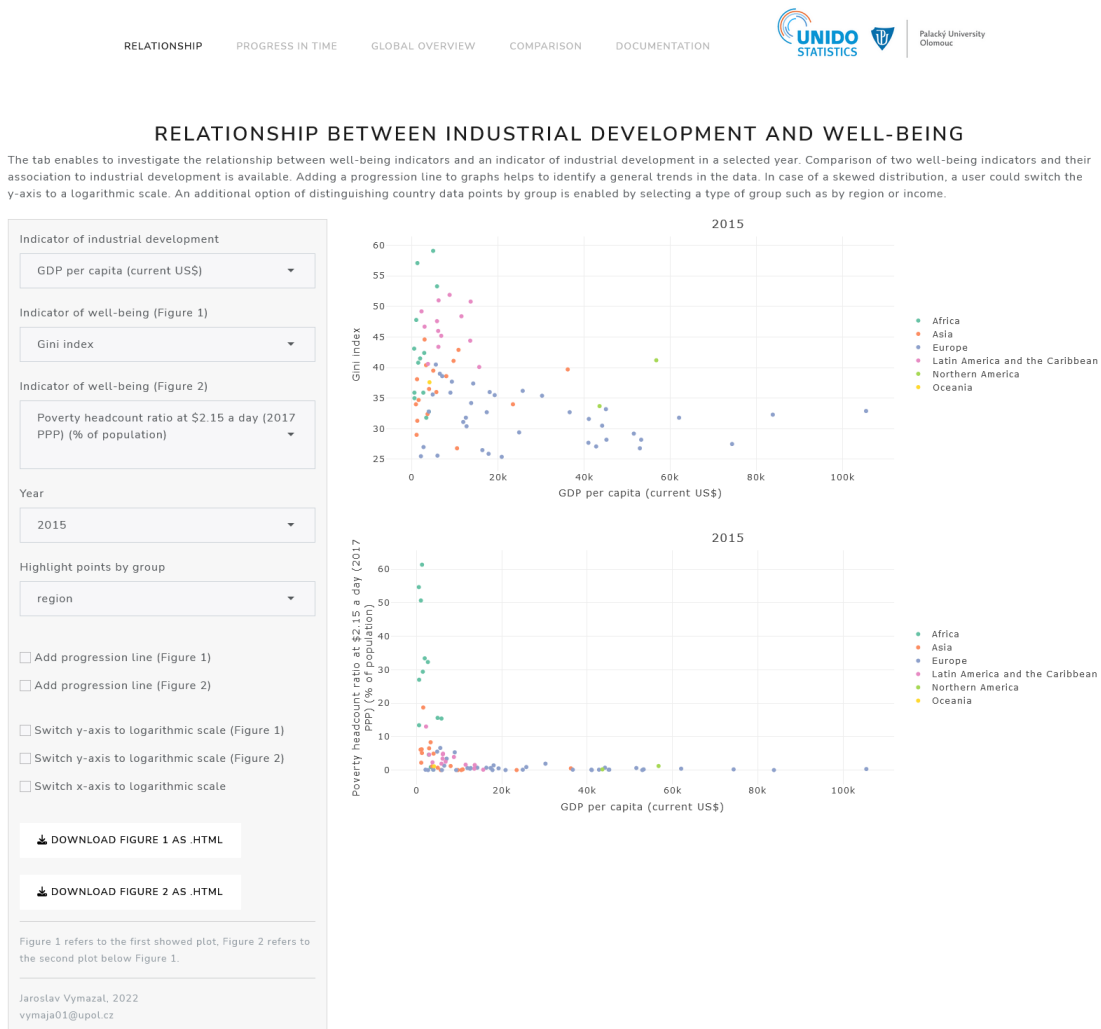
ACKNOWLEDGEMENTS

The creation of this application would not be possible without the help of UNIDO and Palacký University Olomouc.

Obrázek 3.1: Záložka Documentation

3.2. Záložka Relationship

Relationship je první ze čtyř záložek sloužících k vizualizaci. V její levé části je okno, kde si uživatel volí hodnoty vstupů, a v pravé části je pak výstup v podobě samotné vizualizace, kterou jsou v tomto případě dva bodové grafy (viz obrázek 3.2).



Obrázek 3.2: Záložka Relationship

Co se týče vstupů, uživatel má možnost zvolit si jaký ukazatel se bude zobrazovat na ose x , resp. y . Osa x je přitom v případě obou grafů stejná a zobrazuje se na ní uživatelem vybraný ukazatel průmyslového rozvoje. Osy y mohou být pro oba

grafy rozdílné a jsou na nich zobrazovány vybrané ukazatele životní úrovně. Díky tomu může uživatel provádět srovnání vztahů dvou různých ukazatelů životní úrovně a vybraného ukazatele průmyslového rozvoje.

Dále je zde také možnost zvolit si rok, pro který chceme hodnoty ukazatelů sledovat. Rok je stejný pro oba grafy, aby bylo případné srovnání grafů relevantní.

Uživatel si pak může také zvolit jakou barvu budou mít body grafu, jestli chce v grafu zobrazit regresní křivku nebo případně změnit měřítko na jedné z os ze standardního lineárního na logaritmické. Jsou zde i dvě tlačítka (každé pro jeden graf), které uživateli dovolují stáhnout si vytvořené vizualizace v .html formátu, aby byla zachována jejich interaktivita. Pokud se uživatel spokojí i se statickým obrázkem, je možné grafy stáhnout i ve formátu .png pomocí tlačítka nacházejícího se přímo v grafu. Celé okno se vstupy je vidět na obrázku 3.3.

Indicator of industrial development
GDP per capita (current US\$)

Indicator of well-being (Figure 1)
Gini index

Indicator of well-being (Figure 2)
Poverty headcount ratio at \$2.15 a day (2017 PPP) (% of population)

Year
2015

Highlight points by group
region

Add progression line (Figure 1)
 Add progression line (Figure 2)
 Switch y-axis to logarithmic scale (Figure 1)
 Switch y-axis to logarithmic scale (Figure 2)
 Switch x-axis to logarithmic scale

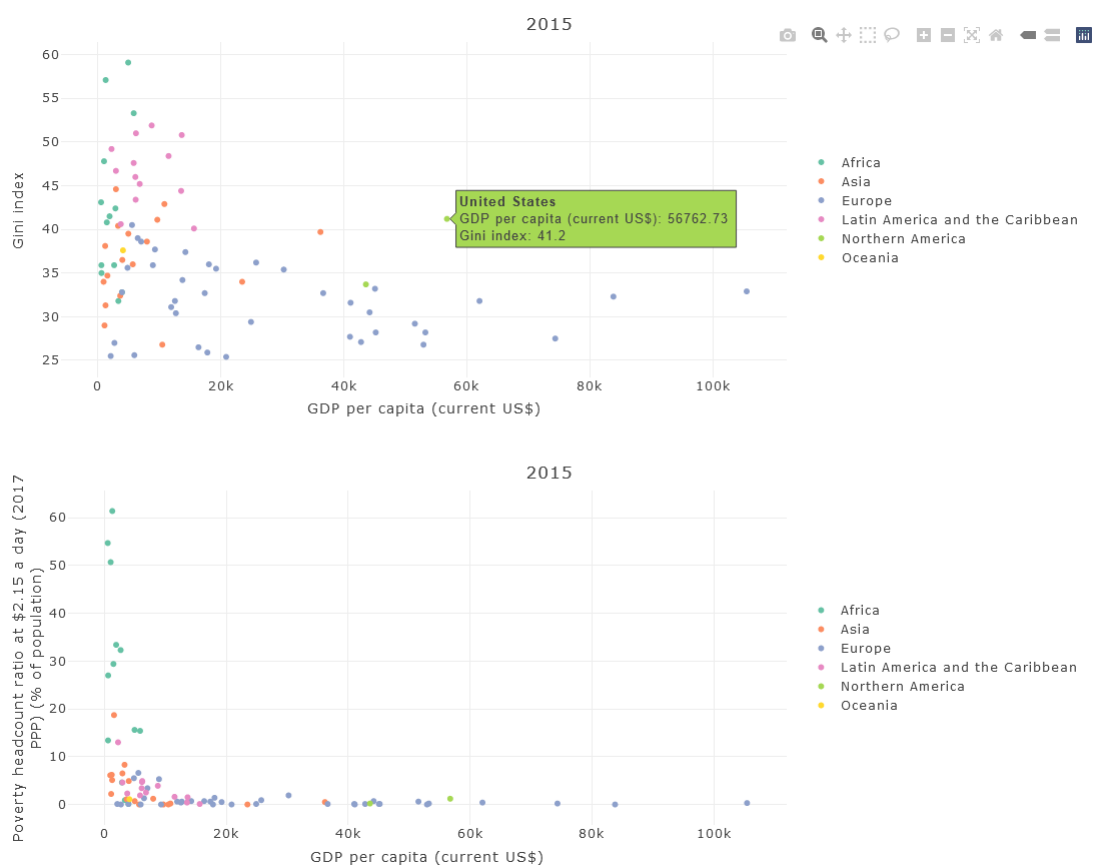
[DOWNLOAD FIGURE 1 AS .HTML](#)
[DOWNLOAD FIGURE 2 AS .HTML](#)

Figure 1 refers to the first showed plot, Figure 2 refers to the second plot below Figure 1.

Jaroslav Vymazal, 2022
vymaja01@upol.cz

Obrázek 3.3: Okno se vstupy v levé části záložky Relationship

V pravé části záložky se zobrazují vizualizace, v tomto případě dva jednoduché bodové grafy. Jelikož jsou grafy interaktivní, uživatel si je může různě přibližovat nebo oddalovat, může si zvolit zobrazení jen určitých skupin bodů pomocí kliknutí na příslušné prvky legendy, a v případě, že najede kurzorem myši na některý z bodů, zobrazí se okno s popisky obsahující jméno daného státu a jemu příslušné hodnoty sledovaných ukazatelů (viz obrázek 3.4).

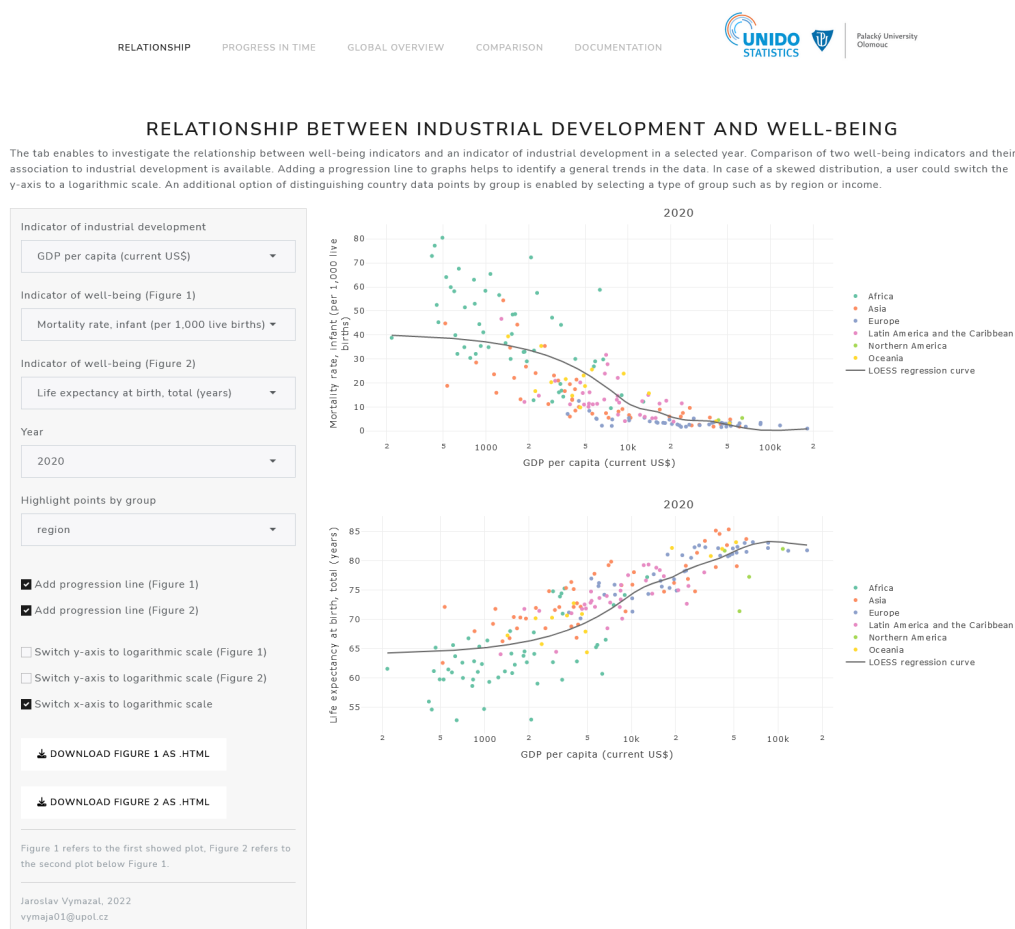


Obrázek 3.4: Vizualizace ze záložky Relationship

Příklad použití této záložky je vidět na obrázku 3.5. Na ose x zobrazujeme na obou grafech hodnoty HDP na obyvatele a na ose y máme v horním grafu úmrtnost novorozenců a ve spodním střední délku života při narození. Jelikož rozdělení HDP na obyvatele je sešikmené, změním měřítko osy x z lineárního na logaritmické. Dále zvolíme rok 2020, jelikož je to poslední rok, pro který je

dostupné reprezentativní množství dat. Body grafů pak obarvíme podle kontinentální příslušnosti jednotlivých zemí a necháme si v obou grafech vykreslit i regresní křivky pro lepší představu o vztahu ukazatelů.

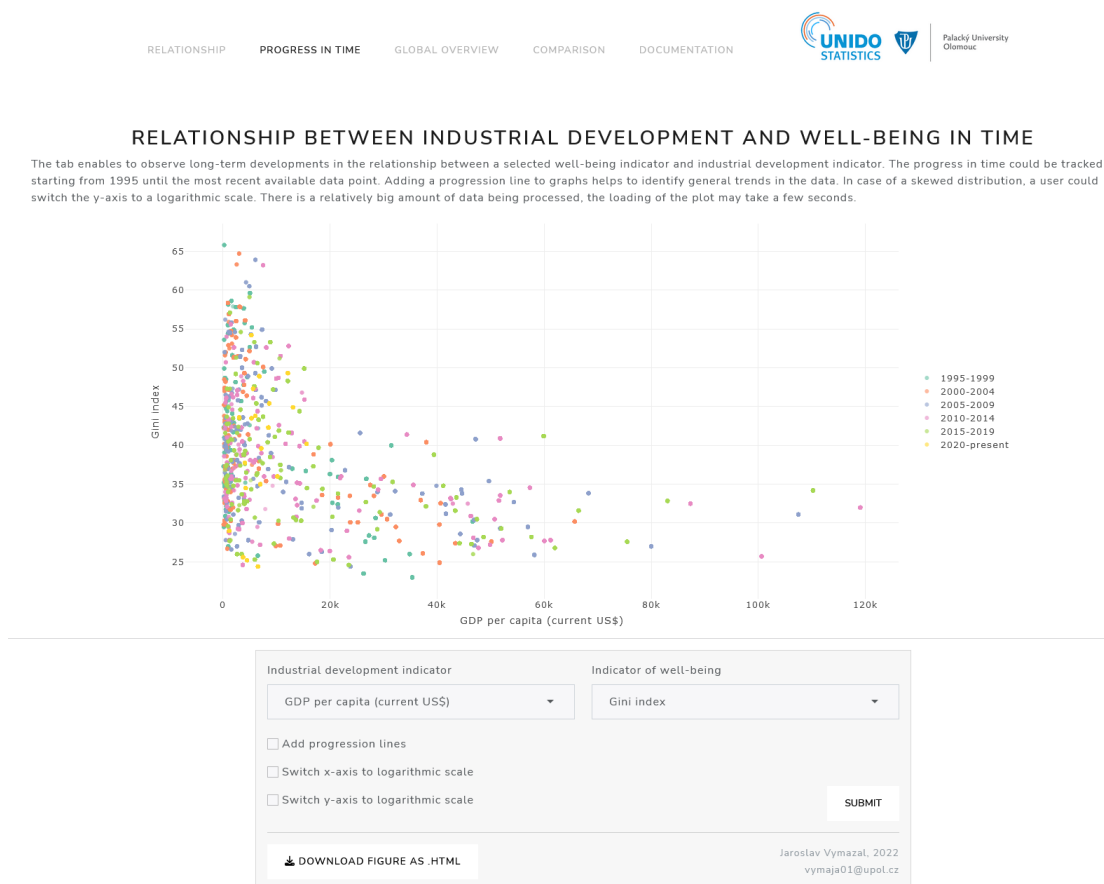
Z horního grafu můžeme vyčíst, že v zemích s vyšší hodnotou HDP na obyvatele bývá úmrtnost novorozenců výrazně nižší než v zemích, kde je hodnota tohoto ukazatele nízká. Na dolním grafu pozorujeme opačný trend. Střední délka života při narození bývá vyšší v zemích s vyšší hodnotou HDP na obyvatele. Z obarvení bodů pak lze vyčíst, že nejnižší hodnoty střední délky života při narození a nejvyšší úmrtnost novorozenců můžeme spatřovat v Africe a některých zemích Asie a Oceánie, kde jsou hodnoty HDP na osobu nízké.



Obrázek 3.5: Příklad použití záložky Relationship

3.3. Záložka Progress in Time

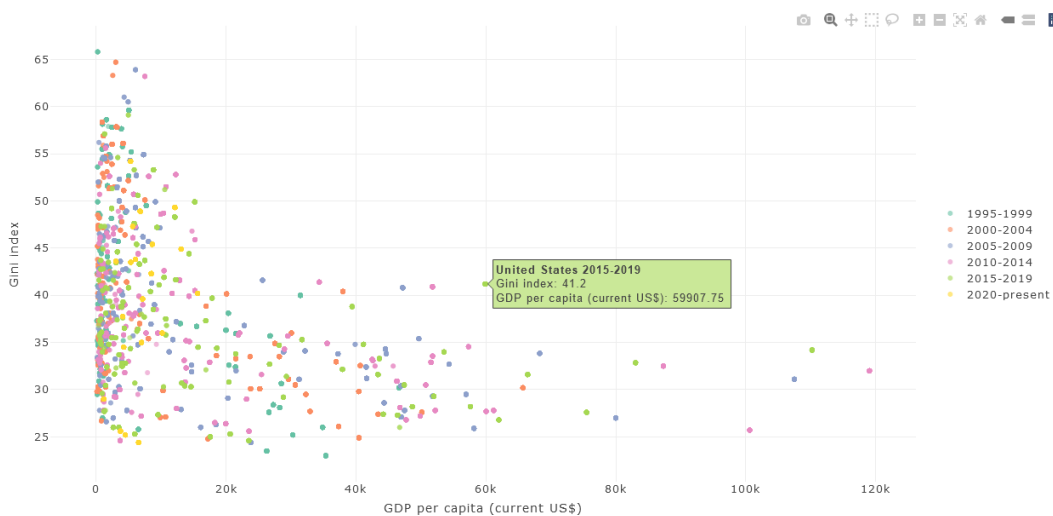
Druhou záložkou sloužící k vizualizaci je Progress in Time. V její horní části je graf, v tomto případě opět bodový graf, a v dolní části se nachází okno, kde si uživatel volí hodnoty vstupů (viz obrázek 3.6).



Obrázek 3.6: Záložka Progress in Time

Zobrazovaný graf v tomto případě slouží k sledování vývoje vztahu dvou ukazatelů v čase. Body grafu jsou obarveny podle časového intervalu, do kterého spadají, přičemž při tvorbě grafu uvažujeme data od roku 1995 do současnosti a jednotlivé intervaly jsou tvořeny pětiletými časovými úseky. Hodnoty ukazatelů pro určitý časový úsek získáváme jako medián všech hodnot daného ukazatele v tomto časovém úseku (vždy pro určitý stát). Stejně jako body jsou obarveny i

případné regresní křivky, které si uživatel může v grafu zobrazit. Každá z křivek v takovém případě odpovídá jednomu z časových intervalů. Graf je opět interaktivní, tzn. uživatel si ho může různě přibližovat nebo případně oddalovat, může si kliknutím na příslušné prvky legendy zobrazit data jen pro vybrané časové intervaly a pokud najede kurzorem myši na některý z bodů, zobrazí se okno s popisky (viz obrázek 3.7).



Obrázek 3.7: Vizualizace ze záložky Progress in Time

Pod grafickým výstupem se nachází okno, kde si uživate může zvolit jaký ukazatel se bude zobrazovat na ose x , resp. y , a jaké měřítko bude pro kterou osu použito (lineární nebo logaritmické). Opět je zde také možnost přidat do grafu regresní křivky a tlačítko pro stáhnutí grafu ve formátu .html. Graf je opět možné stáhnout i jako statický obrázek ve formátu .png pomocí tlačítka přímo v grafu. Narozdíl od předchozí záložky zde vizualizace nereaguje na změny vstupů automaticky, ale pro provedení změn je třeba zmáčknout tlačítko Submit v pravém dolním rohu okna (viz obrázek 3.8).

Industrial development indicator

GDP per capita (current US\$)

Indicator of well-being

Gini index

Add progression lines

Switch x-axis to logarithmic scale

Switch y-axis to logarithmic scale

SUBMIT

DOWNLOAD FIGURE AS .HTML

Jaroslav Vymazal, 2022
vymaja01@upol.cz

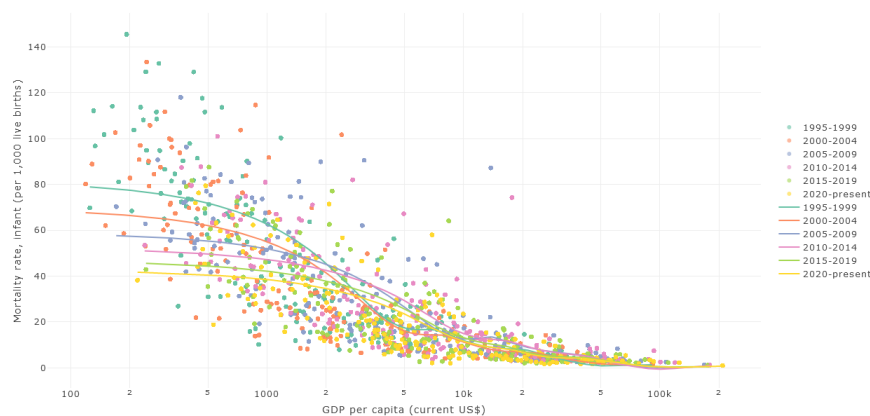
Obrázek 3.8: Okno se vstupy v dolní části záložky Progress in Time

Na obrázku 3.9 je vidět příklad použití této záložky. Sledujeme vývoj vztahu úmrtnosti novorozenců a HDP na obyvatele. Na ose x zobrazujeme HDP na obyvatele a na ose y úmrtnost novorozenců. Kvůli šikmosti rozdělení HDP na obyvatele je na ose x opět použito logaritmické měřítko. Dále jsou v grafu zobrazené regresní křivky pro lepší představu o vývoji vztahu v čase.

Z grafu je patrné, že charakter vztahu těchto dvou ukazatelů se v průběhu let nezměnil, jelikož pro každý časový úsek platí, že se zvyšujícím se HDP na obyvatele se snižuje úmrtnost novorozenců. V průběhu let však lze pozorovat výrazné snížení úmrtnosti novorozenců v zemích s nižšími hodnotami HDP na obyvatele. To vede ke zmenšování rozdílu v hodnotách tohoto ukazatele mezi zeměmi s nižšími a vyššími hodnotami HDP na obyvatele, což se projevuje postupným „zplošťováním“ regresních křivek v čase.

RELATIONSHIP BETWEEN INDUSTRIAL DEVELOPMENT AND WELL-BEING IN TIME

The tab enables to observe long-term developments in the relationship between a selected well-being indicator and industrial development indicator. The progress in time could be tracked starting from 1995 until the most recent available data point. Adding a progression line to graphs helps to identify general trends in the data. In case of a skewed distribution, a user could switch the y-axis to a logarithmic scale. There is a relatively big amount of data being processed, the loading of the plot may take a few seconds.



Industrial development indicator

GDP per capita (current US\$) ▼

Indicator of well-being

Mortality rate, infant (per 1,000 live births) ▼

Add progression lines

Switch x-axis to logarithmic scale

Switch y-axis to logarithmic scale

SUBMIT

[↓ DOWNLOAD FIGURE AS .HTML](#)

Jaroslav Vymazal, 2022
vymaja01@upol.cz

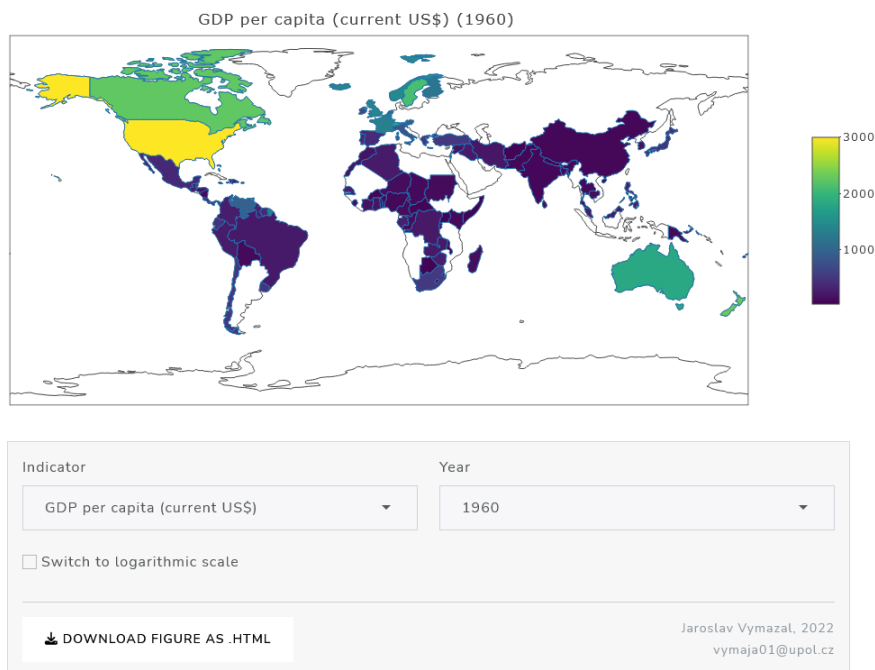
Obrázek 3.9: Příklad použití záložky Progress in Time

3.4. Záložka Global Overview

Záložka Global Overview slouží k vykreslování map. Její rozložení je velmi podobné záložce Progress in Time, tzn. v horní části se nachází samotná mapa a v dolní části je okno, kde si uživatel volí vstupy (viz obrázek 3.10).

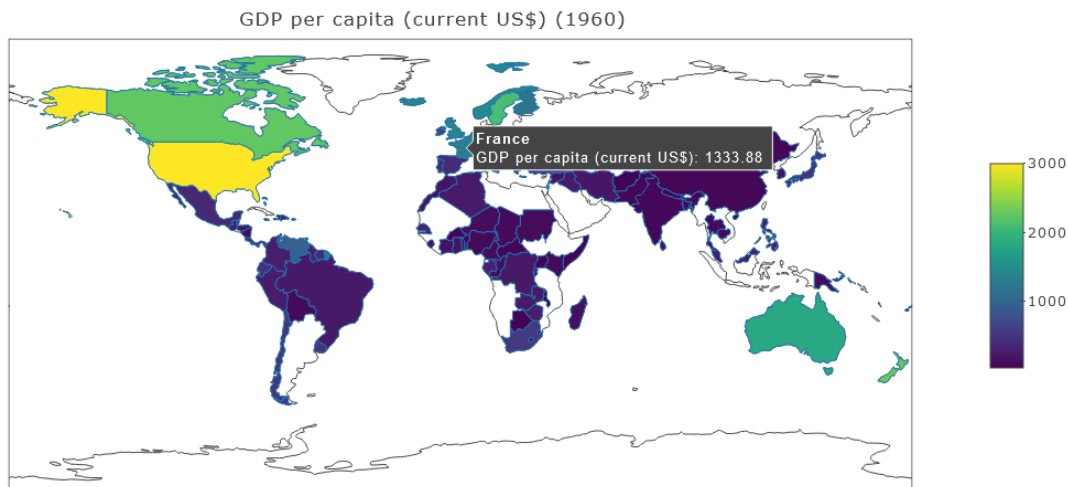
GLOBAL OVERVIEW OF SELECTED INDICATORS

The tab provides information about the performance of a selected indicator by country in a selected year in the global context. The world map enables to identify where the major data gaps appear. In case of a skewed distribution, a user could switch to a logarithmic scale.



Obrázek 3.10: Záložka Global Overview

Mapy zobrazující se v horní části okna slouží k porovnávání hodnoty vybraného ukazatele mezi státy a stejně jako předchozí grafy, je interaktivní. Uživatel si ji může libovolně přibližovat, oddalovat nebo posouvat a v případě, že najede kurzorem myši na některý ze států, zobrazí se okno s popisem, kde je jméno státu a jemu příslušná hodnota zkoumaného ukazatele (viz obrázek 3.11).



Obrázek 3.11: Vizualizace ze záložky Global Overview

V okně pod grafickým výstupem si uživatel může zvolit, z kterého roku budou zobrazovaná data a jaký z ukazatelů se bude na mapě zobrazovat. Opět je zde také možnost změnit měřítko mapy z lineárního na logaritmické a nechybí ani tlačítko pomocí kterého si uživatel může stáhnout mapu ve formátu .html, aby byla zachována její interaktivita. Přímou v mapě je pak také tlačítko umožňující její stažení ve statickém formátu .png. Celé okno je vidět na obrázku 3.12.

Indicator	Year
<div style="border: 1px solid #ccc; padding: 2px; display: flex; justify-content: space-between; align-items: center;"> GDP per capita (current US\$) ▼ </div>	<div style="border: 1px solid #ccc; padding: 2px; display: flex; justify-content: space-between; align-items: center;"> 1960 ▼ </div>
<input type="checkbox"/> Switch to logarithmic scale	
<div style="border: 1px solid #ccc; padding: 5px; display: inline-block; margin-bottom: 5px;"> 📄 DOWNLOAD FIGURE AS .HTML </div>	
<small>Jaroslav Vymazal, 2022 vymaja01@upol.cz</small>	

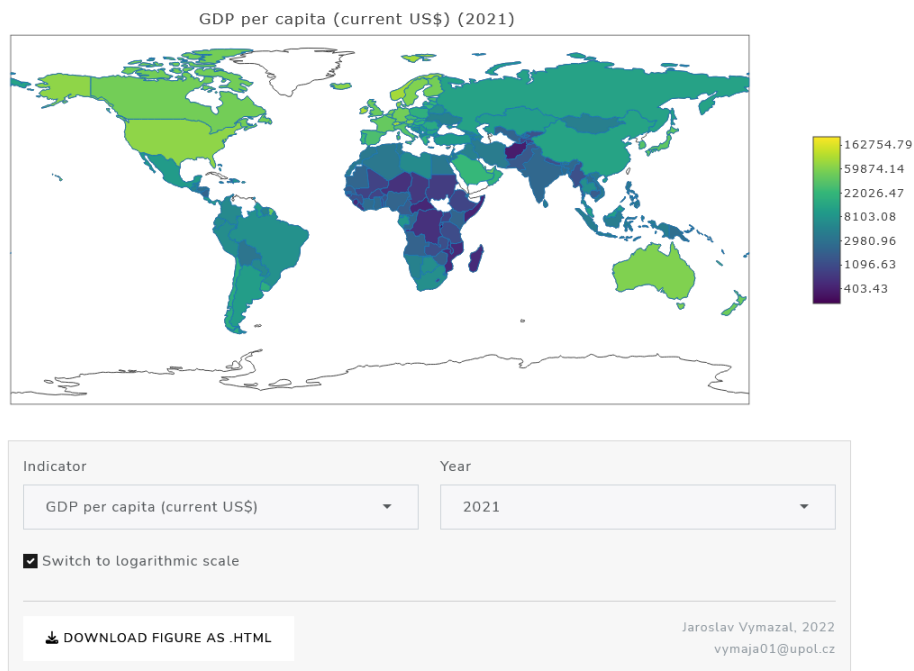
Obrázek 3.12: Okno se vstupy v dolní části záložky Global Overview

Na obrázku 3.13 je vidět příklad použití záložky Global Overview. Jako sledovaný ukazatel jsme zvolili HDP na obyvatele a zajímají nás data z roku 2021. Vzhledem k tomu, že rozdělení sledovaného ukazatele bývá typicky sešikmené, zvolili jsme logaritmické měřítko místo lineárního.

Vidíme, že HDP na obyvatele nabývá nejnižších hodnot v afrických zemích a ve střední a jižní Asii. Nejvyšší hodnoty pak můžeme pozorovat v Severní Americe, Austrálii, Evropě, Japonsku a Jižní Korei.

GLOBAL OVERVIEW OF SELECTED INDICATORS

The tab provides information about the performance of a selected indicator by country in a selected year in the global context. The world map enables to identify where the major data gaps appear. In case of a skewed distribution, a user could switch to a logarithmic scale.



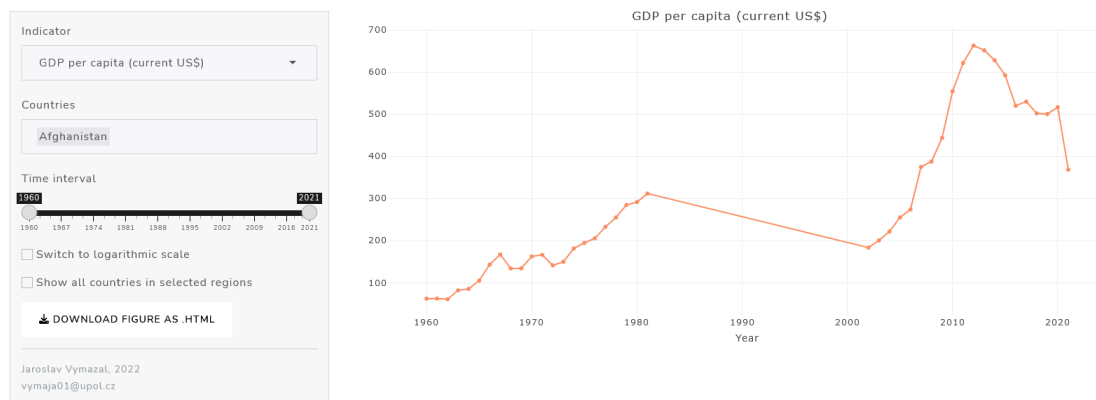
Obrázek 3.13: Příklad použití záložky Global Overview

3.5. Záložka Comparison

Comparison je poslední ze záložek sloužících k vizualizaci. Má podobné rozložení jako záložka Relationship, tzn. na pravé straně se nachází okno se vstupy a na levé výstup v podobě grafu, kterým je v tomto případě spojnicový graf (viz obrázek 3.14).

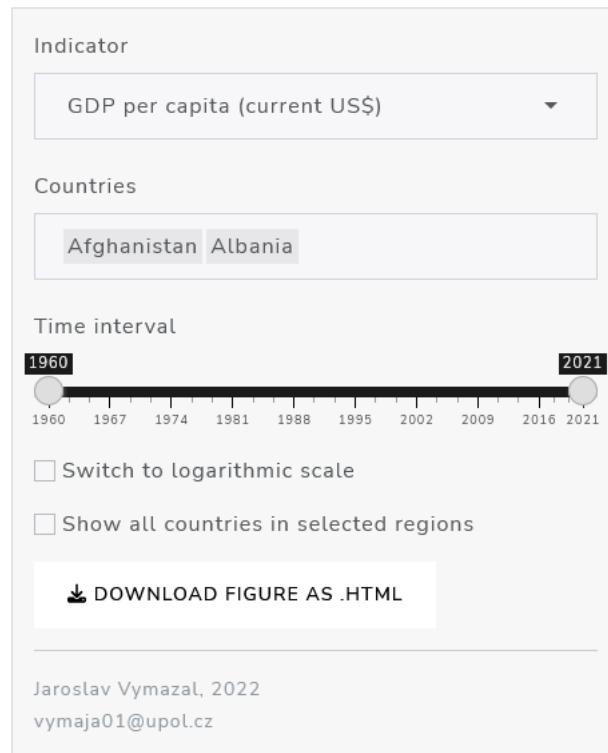
TRENDS IN SELECTED INDICATORS BY COUNTRY

The tab enables the user to investigate a selected indicator and compare its values for one or more countries over the time. The user has an opportunity to adjust the timespan and add data for all other countries in the same region in the background for regional comparison.



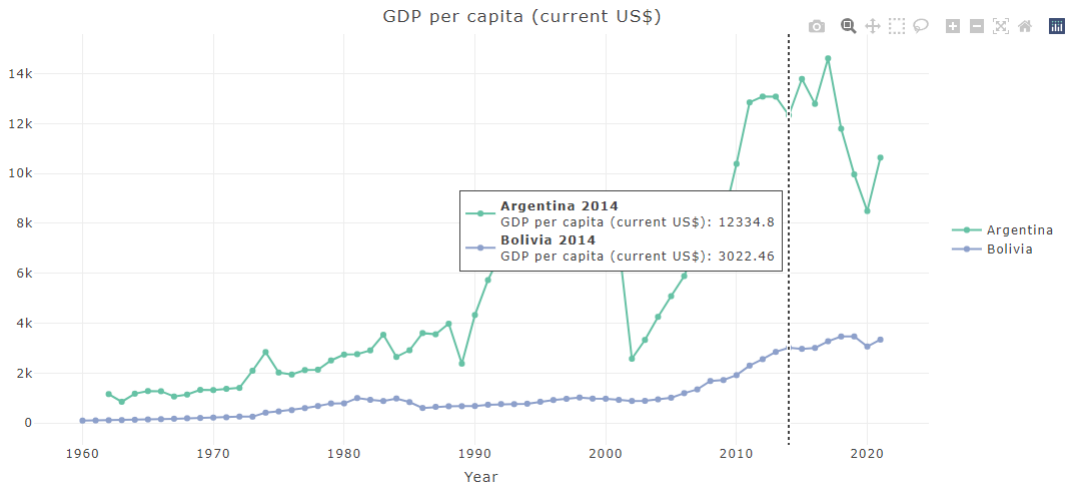
Obrázek 3.14: Zálložka Comparison

Co se týče okna se vstupy, uživatel si zde může zvolit sledovaný ukazatel a časové období, pro které budou data o zvoleném ukazateli sledována. Dále si může vybrat které státy mezi sebou chce porovnávat, přičemž zde není žádné omezení na jejich počet. Uživatel si může vybrat i pouze jeden stát, pokud chce např. sledovat trend vývoje určitého ukazatele v tomto státě a nechce provádět srovnání s ostatními státy. V případě, že by uživatele zajímalo jak si stát vede v rámci svého regionu, může si nechat v pozadí zobrazit všechny státy z tohoto regionu pomocí příslušného tlačítka v okně. Opět je zde také možnost změnit měřítko osy y z lineárního na logaritmické, tlačítko pro stáhnutí grafu ve formátu .html, které se nachází přímo v okně, a tlačítko pro stáhnutí grafu jako statického obrázku ve formátu .png, které je umístěno v samotném grafu. Celé okno se vstupy je vidět na obrázku 3.15.

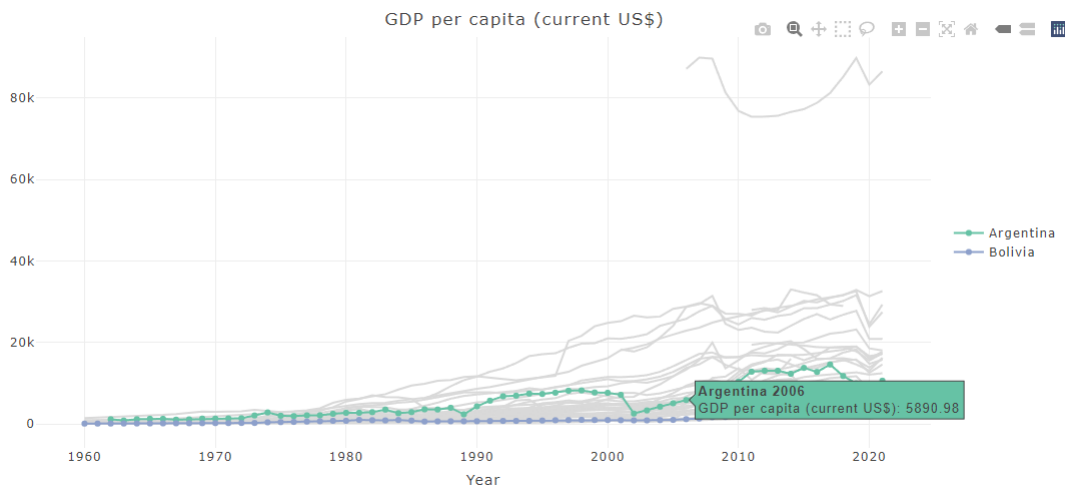


Obrázek 3.15: Okno se vstupy v levé části záložky Comparison

Samotný spojnicový graf je pak opět interaktivní, tzn. uživatel si ho může libovolně přibližovat nebo oddalovat a v případě, že najede na některý z bodů v grafu, zobrazí se okno s popisky obsahující hodnoty sledovaného ukazatele v daném roce pro všechny vybrané státy (viz obrázek 3.16). Toto okno však vypadá odlišně v případě, že si uživatel zvolí zobrazení křivek všech zemí z příslušného regionu v pozadí (viz obrázek 3.17).



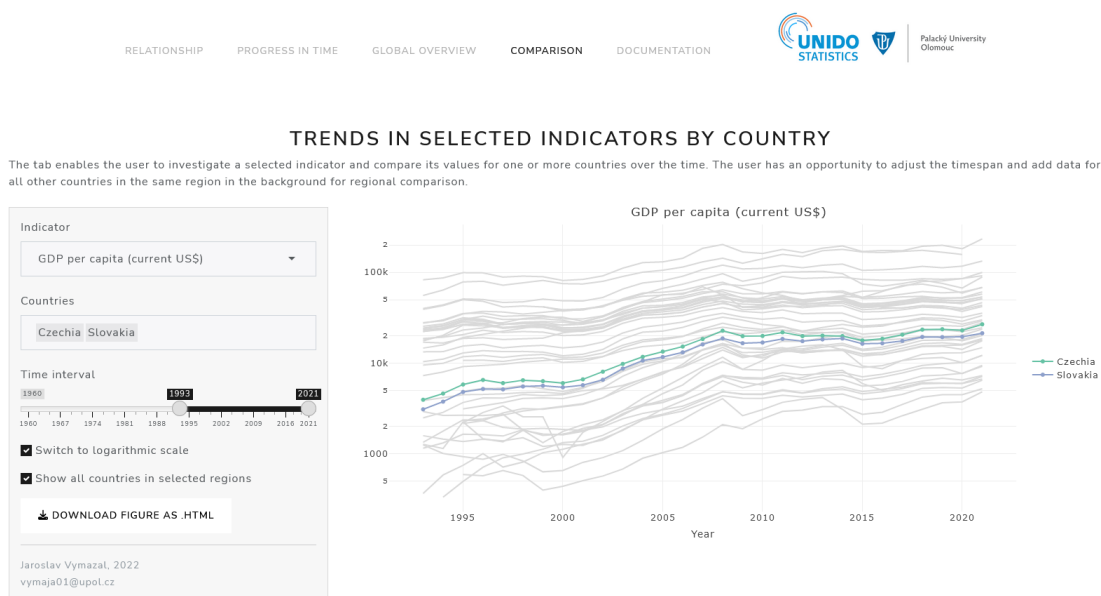
Obrázek 3.16: Vizualizace ze záložky Comparison bez křivek v pozadí



Obrázek 3.17: Vizualizace ze záložky Comparison s křivkami v pozadí

Příklad použití záložky Comparison je vidět na obrázku 3.18. Jako sledovaný ukazatel bylo zvoleno HDP na osobu a porovnávat mezi sebou chceme Česko a Slovensko, přičemž zajímat nás budou data z let 1993 až 2021. Opět, vzhledem k charakteru HDP na osobu, bylo místo lineárního měřítka zvoleno logaritmické. Zvolená je i možnost zobrazení křivek států ve stejném regionu (tj. v tomto případě v Evropě) na pozadí, jelikož chceme porovnání dvou vybraných států chápat v kontextu celé Evropy.

Z výsledného grafu je patrné, že HDP na osobu v Česku a na Slovensku má velmi podobný rostoucí trend, přičemž v Česku je jeho hodnota dlouhodobě vyšší než na Slovensku. V kontextu celé Evropy je vidět, že ani u jednoho ze sledovaných států není hodnota HDP nijak zvlášť vysoká a oba státy se v posledních letech pohybují okolo mediánu hodnot HDP na osobu v Evropě.



Obrázek 3.18: Příklad použití záložky Comparison

Závěr

Výsledkem této práce je aplikace sloužící k vizualizaci dat o ekonomických ukazatelech. Aplikace byla vytvořena v softwaru R, přičemž hlavní roli při její tvorbě hrály balíčky Shiny, umožňující tvorbu webových aplikací v R, a Plotly, díky kterému jsou všechny grafy v aplikaci interaktivní. Dále aplikace využívá API Světové banky, z jejíž databáze si tímto způsobem bere data potřebná k provádění vizualizací.

Aplikace byla vytvářena jako analytický nástroj pro UNIDO a věřím, že jejím uživatelům bude přínosem a usnadní jim práci s daty. Jsem rád, že jsem v rámci této bakalářské práce mohl pracovat na něčem praktickém a zdokonalit se v práci se softwarem R a v programování obecně, což považuji za velmi přínosné.

Literatura

- [1] Databáze Světové banky [online]. [cit. 2023-02-24]. Dostupné z: <https://data.worldbank.org/>.
- [2] Giniho index - Wikipedie [online]. [cit. 2023-02-24]. Dostupné z: https://cs.wikipedia.org/wiki/Giniho_koeficient.
- [3] Jurečka, V., Hlaváček, K., Jánošíková, I., Kolcunová, E., Macháček, M., Paličková, I., Spáčilová, L., Šimek, M.: *Úvod do ekonomie*, 4. aktualizované vydání. VŠB - Technická univerzita Ostrava, Ostrava, 2015.
- [4] What is manufacturing value added? - UNIDO [online]. [cit. 2023-02-28]. Dostupné z: <https://stat.unido.org/content/learning-center/what-is-manufacturing-value-added%253f>.
- [5] Fačevicová, K., Kynčlová, P., Upadhyaya, S.: *How Industrial Development Matters to the Well-being of the Population: Some Statistical Evidence*. UNIDO, 2020.
- [6] Hastie, T., Tibshirani, R., Friedman, J. H.: *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, 2. vydání. Springer, 2009.
- [7] Weisberg, S., Fox, J.: *An R Companion to Applied Regression*, 2. vydání. Thousand Oaks: Sage, 2011.
- [8] International Standard Industrial Classification of All Economic Activities (ISIC), Rev. 4 [online]. [cit. 2023-03-15]. Dostupné z: https://unstats.un.org/unsd/publication/seriesm/seriesm_4rev4e.pdf
- [9] Standard International Trade Classification, Revision 4 [online]. [cit. 2023-03-15]. Dostupné z: https://unstats.un.org/unsd/publication/SeriesM/SeriesM_34rev4E.pdf
- [10] Gini Coefficient and Lorenz Curve Around the World [online]. [cit. 2023-03-25]. Dostupné z: <https://knoema.com/crmndag/gini-coefficient-and-lorenz-curve-around-the-world>