EKONOMICKÁ UNIVERZITA V BRATISLAVE

FAKULTA HOSPODÁRSKEJ INFORMATIKY

Evidenčné číslo: 103005/I/2022/36080377336913924

Analýza dát v programovacom jazyku Python

Diplomová práca

2022 **Bc. Matúš Kudláč**

EKONOMICKÁ UNIVERZITA V BRATISLAVE

FAKULTA HOSPODÁRSKEJ INFORMATIKY

Analýza dát v programovacom jazyku Python

Diplomová práca

Študijný program: informačný manažment
Študijný odbor: ekonómia a manažment
Školiace pracovisko: Katedra štatistiky FHI
Vedúci záverečnej práce: Ing. Silvia Komara, PhD.

Bratislava 2022 **Bc. Matúš Kudláč**

Ekonomická univerzita v Bratislave
Fakulta hospodárskej informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Bc. Matúš Kudláč
Študijný program: informačný manažment (Jednoodborové štúdium, inžiniersky II. st., denná forma)
Študijný odbor: ekonómia a manažment
Typ záverečnej práce: Inžinierska záverečná práca
Jazyk záverečnej práce: slovenský
Sekundárny jazyk: anglický
Názov: Analýza dát v programovacom jazyku Python

Anotácia: Python je moderný programocí jazyk, ktorého popularita stále rastie. Cieľom práca bude použitie základných štatistických metód a analýza vybraných dát v tomto jazyku, so zameraním na pokročilé grafické zobrazenia.

Vedúci: Ing. Silvia Komara, PhD. **Katedra:** KŠ FHI - Katedra štatistiky FHI **Vedúci katedry:** doc. Ing. Mária Vojtková, PhD.

Dátum zadania: 03.11.2020

Dátum schválenia: 05.11.2020 doc. Ing. Mária Vojtková, PhD.
vedúci katedry

Poďakovanie

Touto cestou by som sa chcel poďakovať vedúcej diplomovej práce, Ing. Silvii Komara, PhD., za pomoc, odborné vedenie, cenné rady a pripomienky pri vypracovaní mojej diplomovej práce.

Abstrakt

KUDLÁČ, Matúš: *Analýza dát v programovacom jazyku Python.* – Ekonomická univerzita v Bratislave. Fakulta hospodárskej informatiky; Katedra štatistiky. – Vedúci záverečnej práce: Ing. Silvia Komara, PhD. Bratislava: FHI, 2022, 50 strán.

Cieľom záverečnej práce je ukázať možnosti programovacieho jazyka Python v oblasti vizualizácie dát a prezentovať vybrané knižnice daného jazyka, ktoré je k tomu možné použiť. Práca je rozdelená do 5 kapitol. Obsahuje 10 obrázkov, 10 tabuliek a 8 grafov. V prvej kapitole je uvedená základná definícia analýzy dát, charakteristika programovacieho jazyka Python a popis jeho histórie, ako aj opis knižníc použitých v práci a takisto aj opis Python distribúcie Anaconda a vývojového prostredia Jupyter. Druhá kapitola sa zameriava na bližší opis cieľov danej práce. V tretej kapitole je uvedená všeobecná charakteristika použitých typov grafov a štatistických ukazovateľov spoločne so zoznamom použitých knižníc a ich konkrétnych verzií. Vo štvrtej kapitole sú zobrazené jednotlivé grafy spolu so zdrojovým kódom použitým na ich zobrazenie a takisto je tu uvedený aj zdroj a popis jednotlivých zdrojových dát, ktoré sú v práci použité. Záverečná kapitola sa venuje diskusii k výsledkom danej práce. Výsledkom riešenia danej problematiky je grafické znázornenie skúmaných údajov a vzťahov medzi nimi spolu so zdrojovými kódmi v programovacom jazyku Python použitými na ich zobrazenie.

Kľúčové slová: Python, Jupyter, Anaconda, graf, dáta, štatistika, analýza

Abstract

KUDLÁČ, Matúš: *Data analysis in the Python programming language.* – University of Economics in Bratislava. Faculty of Economic informatics; Department of Statistics. – Thesis supervisor: Ing. Silvia Komara, PhD. Bratislava: FHI, 2022, 50 pages.

The aim of the thesis is to show the possibilities of the Python programming language in the field of data visualization and to present the selected libraries of the given language that can be used for that. The thesis is divided into 5 chapters. It contains 10 pictures, 10 tables and 8 graphs. The first chapter provides a basic definition of the data analysis, characteristics of the Python programming language and a description of its history, as well as a description of the libraries used in the thesis and also a description of the Python distribution Ananconda and the Jupyter development environment. The second chapter focuses on a more detailed description of the objectives of the thesis. The third chapter presents the general characteristics of the types of graphs and statistical indicators used, together with a list of the libraries used and their specific versions. In the fourth chapter, the created graphs are displayed together with the source code that has been used to display them and also the source and description of the source data. The final chapter is devoted to the discussion of the results of the work. The result of the thesis is a graphical representation of the examined data and the relationships between them, together with the source code in the Python programming language used to display them.

Keywords: Python, Jupyter, Anaconda, graph, data, statistics, analysis

Obsah

Zoznam obrázkov, tabuliek a grafov

Úvod

Metódy získavania, spracovania a prezentácie dát sa postupne menili vplyvom ľudského rozvoja a získavania nových poznatkov. Spolu s tým bolo potrebné zameriavať sa aj na čoraz novšie spôsoby ich spracovania a zobrazovania. V posledných rokoch sa vplyvom informačného rozvoja dostupnosť a množstvo dostupných dát zvyšujú vysokým tempom. Z tohto dôvodu je nevyhnutné tieto dáta spracovávať s použitím informačných technológií a vyvíjať nové technológie, ktoré umožnia spracovávať tieto dáta rýchlo a efektívne s možnosťou ich prehľadnej a pochopiteľnej vizualizácie.

Programovací jazyk Python si s postupom času získava čoraz väčšiu popularitu z dôvodu jednoduchého použitia daného jazyka na rôznych platformách a takisto aj veľkého množstva dostupných knižníc pokrývajúcich veľké množstvo oblastí, pričom tento jazyk sa neustále vyvíja a prispôsobuje pre uspokojenie neustále sa meniacich spoločenských potrieb. Takisto v rámci oblasti analýzy a vizualizácie dát ponúka tento programovací jazyk široké možnosti, pričom je dostupné čoraz väčšie množstvo knižníc daného jazyka zaoberajúcich sa danou problematikou. Z tohto dôvodu si tento programovací jazyk získava čoraz väčšiu popularitu aj v tejto oblasti.

V tejto práci sa zameriame na skúmanie a prezentáciu možností programovacieho jazyka Python v oblasti vizualizácie dát, pričom použijeme viaceré dostupné knižnice programovacieho jazyka Python a dáta rôznych typov získané z webovej stránky www.kaggle.com a takisto aj geografické údaje obsahujúce hranice jednotlivých štátov získané z webovej stránky datahub.io. Práca sa zameria na prezentáciu základných, ako aj pokročilejších grafických zobrazení týchto dát, ktoré budú vytvorené pomocou zdrojových kódov napísaných v programovacom jazyku Python s použitím distribúcie tohto programovacieho jazyka Anaconda a vývojového prostredia Jupyter, ktoré použijeme pre vytvorenie zdrojových kódov v programovacom jazyku Python a takisto aj pre zobrazenie výsledkov. Zdrojové kódy budú tvoriť najvýznamnejšiu časť výsledkov danej práce, keďže budú prezentovať možnosti použitých knižníc programovacieho jazyka Python a spôsob, akým sa dajú použiť pre zobrazenie rôznych typov grafov.

Vytvorené grafické zobrazenia rozdelíme do jednotlivých kapitol podľa typu grafu, ktorý každé z daných grafických zobrazení reprezentuje. Na načítanie dát do pamäte a ich spracovanie bude použitá knižnica *pandas*, ktorá umožňuje načítať dáta zo zdrojov rôzneho typu a pracovať s nimi v tabuľkovej forme pomocou objektu *DataFrame*. Pre vykresľovanie grafov bude použitá knižnica *matplotlib*, ktorá je populárnym nástrojom pre vizualizáciu dát v programovacom jazyku Python so širokými možnosťami podrobného nastavenia jednotlivých grafických zobrazení a takisto aj knižnica *seaborn*, ktorá je postavená na knižnici *matplotlib* a poskytuje funkcie pre jednoduchšie zobrazenie pokročilejších zobrazení prostredníctvom viacerých už predpripravených komplexných funkcií. Takisto bude použitá knižnica *geopandas*, ktorá umožňuje vykresľovať geografické údaje a je založená na knižnici *pandas* a umožňuje pracovať s dátami podobným spôsobom pri pridaní možnosti používať geografické dáta.

1 Súčasný stav problematiky doma a v zahraničí

1.1 Analýza dát

Analýza dát znamená skúmanie dát, hľadanie zmysluplných poznatkov z nich a vyvodzovanie záverov. Hlavným cieľom tohto procesu je zhromažďovať, filtrovať, čistiť, transformovať, popisovať, vizualizovať a komunikovať poznatky získané z týchto dát s cieľom objaviť informácie, ktoré sú dôležité pri rozhodovaní. Proces analýzy dát sa skladá z nasledujúcich fáz [15] :

- zber údajov – nájdenie a zbieranie údajov z dostupných zdrojov
- predspracovanie údajov – filtrovanie údajov, čistenie údajov a ich transformácia do požadovaného formátu
- analýza a hľadanie záverov – skúmanie, popisovanie a vizualizácia údajov a hľadanie záverov, ktoré z nich vyplývajú
- interpretácia výsledkov – pochopenie výsledkov a hľadanie vplyvu každej z premenných na celý systém
- komunikovanie výsledkov – komunikovanie výsledkov vo forme, ktorá je pochopiteľná pre bežných ľudí

1.2 Programovací jazyk Python

Odkedy sa programovací jazyk Python (https://www.python.org) prvýkrát objavil v roku 1991, stal sa jedným z najobľúbenejších interpretovaných programovacích jazykov spolu s jazykmi Perl, Ruby a ďalšími. Jazyky Python a Ruby sa stali ešte viac populárnymi v období okolo roku 2005, pretože poskytovali veľké množstvo frameworkov pre tvorbu webových stránok, ako napríklad framework Django pre jazyk Python alebo Rails pre jazyk Ruby (https://www.ruby-lang.org). [1]

1.2.1 História programovacie jazyka Python

Programovací jazyk Python je nasledovník programovacieho jazyka ABC, ktorý bol inšpirovaný jazykmi ALGOL 68 a SETL. Pôvodne ho vytvoril Guido van Rossum ako vedľajší osobný projekt v neskorých osemdesiatych rokoch počas letných prázdnin. V tom čase pracoval v CWI Centrum Wiskunde & Informatica v Holandsku. Vo februári 1991 publikoval svoj kód na alt.sources, pričom táto verzia bola označená ako verzia 0.9.0. Verzia 1.0 bola vydaná v januári 1994. [1]

Verzia 2.0 programovacieho jazyka Python bola vydaná v roku 2000, verzia 2.7.18, ktorá bola poslednou verziou Pythonu 2 bola vydaná 20. apríla 2020. Od 1. januára 2020 je druhá verzia programovacieho jazyka Python oficiálne nepodporovaná. V podobe 2.7.18 je Python 2 zmrazený a neplánuje sa naďalej vyvíjať. [26]

Verzia programovacieho jazyka Python s označením 3.0 bola vydaná 3. decembra 2008. Python 3.0 (známy aj ako "Python 3000" alebo "Py3k") je nová verzia jazyka, ktorá nie je kompatibilná s verziami radu 2.x. Jazyk je vo veľa aspektoch rovnaký, ale viacero detailov, najmä to, ako fungujú vstavané objekty, ako napríklad slovníky a reťazce, sa oproti verziám radu 2.x výrazne zmenilo a veľa zastaraných funkcií bolo odstránených. [2]

1.2.2 Dátové typy v jazyku Python

Python podporuje viacero základných dátových typov, ako napríklad celé čísla a čísla s pohyblivou desatinnou čiarkou, ale takisto podporuje aj celé čísla neobmedzenej dĺžky a komplexné čísla. Podporuje tiež operácie s textovými reťazcami, pričom na rozdiel od textových reťazcov vyskytujúcich sa vo väčšine programovacích jazykov, textové reťazce v jazyku Python sú nemenným typom, takže operácie, ktoré by inak menili reťazec (napríklad zámena znakov), namiesto toho vracajú nový reťazec. [22]

Jedným zo základných aspektov programovacieho jazyka Python je koncept kolekčných (kontajnerových) typov. Vo všeobecnosti kolekcia je objekt, ktorý obsahuje iné objekty tak, že je možné k nim pristupovať pomocou indexov alebo kľúčov. Kolekcie môžu mať dve základné formy: mapované typy a sekvenčné typy. [22]

Mapované typy sú nezoradené typy implementované v podobe asociatívneho poľa, ktoré mapujú množinu objektov alebo kľúčov na elementy v množine hodnôt podobne ako matematické funkcie. Iným typom kolekcií sú zoradené postupnosti, sekvenčné typy, medzi ktoré patria zoznamy, tuple a textové reťazce. Všetky sekvenčné typy sú indexované pozične (od indexu 0 po dĺžku postupnosti – 1) a všetky okrem textových reťazcov môžu obsahovať objekty ľubovoľného typu (textové reťazce môžu obsahovať iba znaky, ktoré sú v programovacom jazyku Python reprezentované ako jednoznakové reťazce). Reťazce a tuple sú nemenné, zatiaľ čo hodnotu zoznamov je možné meniť, čo znamená, že je možné pridávať, odoberať alebo meniť elementy jednotlivých zoznamov. [22]

1.3 Analýza a vizualizácia dát v jazyku Python

1.3.1 NumPy

Knižnica *NumPy* (https://numpy.org) je už dlhé obdobie základným kameňom numerických výpočtov v Pythone. Poskytuje dátové štruktúry a algoritmy potrebné pre väčšinu vedeckých použití zahŕňajúcimi číselné údaje v programovacom jazyku Python. Okrem iného knižnica *NumPy* obsahuje [1]:

- objekt *ndarray* prestavujúci rýchle a efektívne viacrozmerné pole
- funkcie na vykonávanie výpočtov po prvkoch s poľami alebo matematických operácií medzi poľami
- nástroje na čítanie a zapisovanie datasetov na disk
- operácie lineárnej algebry a nástroje na generovanie náhodných čísel

1.3.2 Pandas

Knižnica *pandas* (https://pandas.pydata.org) poskytuje vysokoúrovňové dátové štruktúry a funkcie, ktoré boli vytvorené pre umožnenie rýchlej a jednoduchej práce s dátami. Od svojho vzniku v roku 2010 prispela táto knižnica k tomu, že sa z jazyka Python stal silný a produktívny nástroj pre potreby analýzy dát. Najdôležitejšie objekty, ktoré sa nachádzajú v knižnici *pandas* sú *DataFrame*, ktorý predstavuje tabuľkovú, stĺpcovo orientovanú dátovú štruktúru a *Series*, jednorozmerný objekt obsahujúci indexy jednotlivých prvkov spolu s prvkami samotnými. [1]

1.3.3 Matplotlib

Matplotlib (https://matplotlib.org) je knižnica jazyka Python pre vytváranie grafov a ďalších dvojrozmerných vizualizácii dát. Pôvodne ju vytvoril John D. Hunter a momentálne je vyvíjaná veľkým tímom vývojárov. Je vytvorená pre vytváranie grafov, ktoré sú vhodné na zverejnenie. [1]

1.3.4 Seaborn

Seaborn (https://seaborn.pydata.org) je open source knižnica, ktorá bola vyvinutá pre programovací jazyk Python. *Seaborn* používa knižnicu *matplotlib* ako základnú knižnicu a ponúka jednoduché, ľahko zrozumiteľné, interaktívne a atraktívne vizualizácie. [15]

1.3.5 GeoPandas

GeoPandas je knižnica jazyka Python pre prácu s vektorovými dátami. Je založená na knižnici *pandas*, pričom pridáva možnosť používať priestorové dáta. Knižnica tiež pridáva funkcionalitu z geografických balíčkov jazyka Python. *GeoPandas* ponúka 2 hlavné dátové objekty – *GeoSeries*, založený na objekte *Series* nachádzajúceho sa v knižnici *pandas* a *GeoDataFrame*, založený na *DataFrame* z knižnice *pandas*. Tieto objekty však pridávajú geometrickú hodnotu pre každý riadok. Možnosť čítania a zápisu je dostupná pre väčšinu dostupných formátov vektorových dát. [16]

1.4 Anaconda

Anaconda (https://www.anaconda.com) je open source distribúcia programovacieho jazyka Python vhodná pre spracovanie veľkého množstva dát, prediktívnu analytiku a vedecké výpočty, ktorá sa zameriava na zjednodušenie správy a nasadenia balíčkov. Anaconda v základnej inštalácii obsahuje samotný Python a takisto sú tu zahrnuté aj základné nástroje potrebné pre prácu s týmto jazykom, ako napríklad najvýznamnejšie knižnice jazyka Python a aplikáciu Jupyter Notebook, ktorá umožňuje písanie, úpravu a spúšťanie zdrojového kódu v programovacom jazyku Python, alebo aj balíčkový systém conda. [16]

1.5 Jupyter

Jupyter Notebook (https://www.jupyter.org) je webová aplikácia, ktorú je možné použiť pre vytvorenie súborov, ktoré obsahujú zdrojový kód programovacieho jazyka, napríklad jazyka Python, iný text, obrázky, matematické vzorce a takisto aj grafy. Tieto súbory bývajú často používané pre vzdelávacie účely alebo na demonštráciu jazyka Python. Je tu možné importovať alebo exportovať zdrojové súbory jazyka Python, špeciálne súbory aplikácie Jupyter Notebook

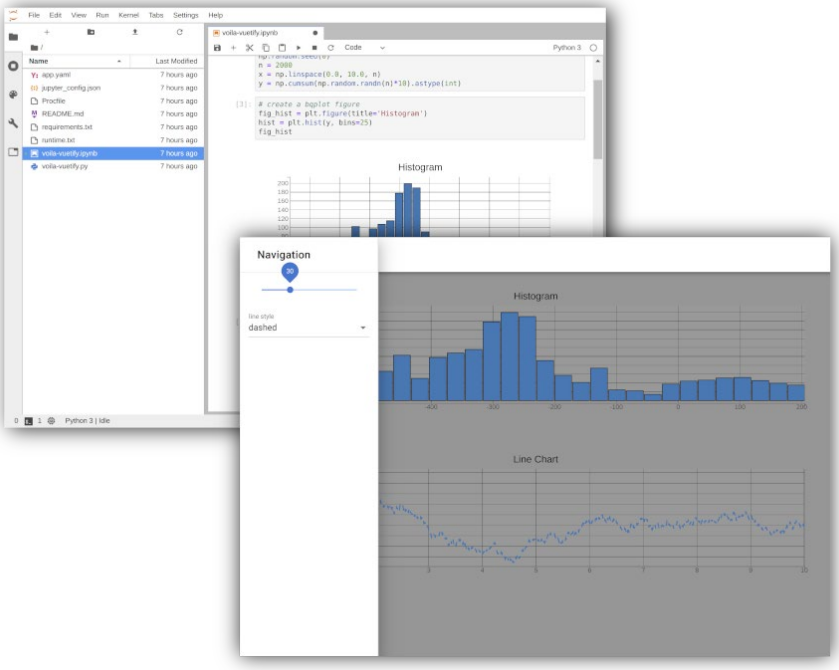
vo formáte .ipynb a takisto môžu byť použité aj viaceré ďalšie formáty súborov. Jupyter môže bežať lokálne, ale takisto môže byť použitý aj online. [15]

Názov Jupyter predstavuje akronym, ktorý znamená Julia, Python a R. Pôvodne boli v tomto prostredí implementované tieto 3 programovacie jazyky, neskôr k nim boli postupne pridané ďalšie programovacie jazyky ako napríklad C, C++, Scala, Perl, Go, PySpark alebo Haskell. [15]

Medzi významné možnosti vývojového prostredia Jupyter patria [15] :

- možnosť editovať zdrojový kód v prehliadači so správnym odsadením
- možnosť spúšťať zdrojový kód priamo v prehliadači
- možnosť zobraziť výstup zdrojového kódu v prehliadači
- možnosť vykresliť grafy, obrázky, videá, tabuľky a ďalšie priamo v prehliadači
- možnosť exportovať zdrojový kód do súborov v rôznych formátoch, ako napríklad PDF, HTML, zdrojový kód jazyka Python, LaTex a ďalšie

Obrázok 1 ***Vývojové prostredie Jupyter***



Zdroj: jupyter.org

1.5.1 História vývojového prostredia Jupyter

Project IPython (https://ipython.org) sa prvýkrát objavil v roku 2001 ako postranný projekt Fernanda Péreza, ktorý sa snažil vytvoriť lepší interaktívny interpreter pre programovací jazyk Python. V nasledujúcich rokoch sa stal jedným z najdôležitejších nástrojov pri práci s týmto jazykom. Aj keď neposkytuje žiadne vlastné výpočtové alebo dátové analytické nástroje, bol vytvorený pre maximalizáciu produktivity v oblasti interaktívnej výpočtovej techniky a vývoja softvéru. Poskytuje tiež jednoduchý prístup k shellu operačného systému a jeho súborovému systému. V roku 2014 predstavili Fernando Pérez spolu s vývojovým tímom IPython projekt Jupyter. [1]

2 Cieľ práce

V práci sa zameriame na základné, ale aj pokročilé grafické zobrazenia prostredníctvom jazyka Python, pričom k tomu použijeme viaceré knižnice jazyka Python, konkrétne *numpy*, *pandas*, *matplotlib*, *seaborn* a *geopandas*. Na tento účel použijeme rôzne typy dát, ktoré sme získali z viacerých dostupných zdrojov, pričom ako primárny zdroj dát bola použitá webová stránka www.kaggle.com, odkiaľ sme získali základné údaje použité pre vizualizáciu, pričom tieto dáta zahŕňajú údaje o niektorých základných charakteristikách svetových krajín na základe informácii pôvodne pochádzajúcich od GHO (Globálne observatórium zdravia) a UNESCO (Organizácia OSN pre vzdelávanie, vedu a kultúru) a takisto informácie o histórii priemerných denných teplôt v jednotlivých mestách.

Spolu s jednotlivými grafickými zobrazeniami budú v práci uvedené zdrojové kódy v programovacom jazyku Python, pomocou ktorých budú jednotlivé zobrazenia vytvorené a ktoré budú najvýznamnejšou súčasťou výsledkov práce, keďže s ich pomocou budú vytvorené jednotlivé zobrazenia a takisto bude možné do budúcna znovu používať a ďalej upravovať tieto zdrojové kódy alebo ich jednotlivé časti pre zobrazenie rovnakých alebo podobných grafických zobrazení.

3 Metodika práce a metódy skúmania

V tejto práci sme použili knižnice programovacieho jazyka Python, ktorými sú *numpy*, *pandas*, *matplotlib*, *seaborn* a *geopandas* a vývojové prostredie Jupyter a takisto aj Anacondu, ktorá je distribúciou jazyka Python. Použité sú údaje z webovej adresy www.kaggle.com, konkrétne údaje o 183 krajinách a ich hodnoty z rokov 2000 až 2016 a priemerné mesačné teploty v Bratislave v rokoch 2012 až 2019. Python je použitý vo verzii 3.9. Na načítanie, uloženie a manipuláciu skúmaných dát je použitá knižnica *pandas* vo verzii 1.4.1. Pre vizualizáciu dát sú použité knižnice *matplotlib* vo verzii 3.5.1 a *seaborn* vo verzii 0.11.2. Na vykreslenie máp je použitá knižnica *geopandas* vo verzii 0.9.0. Knižnica *numpy* bola použitá vo verzii 1.21.5.

3.1 Základné pojmy zo štatistiky

Hromadné javy sú také javy, ktoré sa za presne definovaných podmienok (vecných, časových a priestorových) viackrát vyskytujú, resp. viackrát opakujú, napr. pôrody, úmrtia, nákup spotrebného tovaru, dopravné nehody, výroba automobilov atď. Tieto javy vykazujú určité pravidelnosti (zákonitosti), ktoré možno identifikovať až po ich viacnásobnom opakovaní. Individuálne javy považujeme za konkrétny prejav hromadného javu. [5]

Štatistická jednotka je základný prvok, na ktorom skúmame konkrétny prejav určitého hromadného javu. Štatistickými jednotkami môžu byť akékoľvek objekty: osoby, zvieratá, rastliny, veci, udalosti, organizácie, firmy, predajne, územné jednotky a pod. [5]

Štatistický súbor je množina štatistických jednotiek, ktoré majú požadované spoločné (identické) vlastnosti. Tieto základné spoločné vlastnosti sú podmienkou príslušnosti

štatistických jednotiek do štatistického súboru, a preto sa musia pred každým štatistickým skúmaním presne vymedziť. [5]

Rozsah štatistického súboru je určený počtom jednotiek v súbore. Obyčajne za veľký súbor považujeme súbor s rozsahom väčším ako 30 štatistických jednotiek. Ak počet štatistických jednotiek v súbore nie je väčší ako 30, hovoríme o tzv. malých súboroch. [5]

3.1.1 Priemery

Aritmetický priemer (označujeme ho $\bar{x}$) je definovaný ako podiel súčtu všetkých hodnôt znaku x_i a rozsahu štatistického súboru n. Vypočítame ho pomocou vzťahu [5] :

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{3.1}$$

v ktorom

x_i je hodnota znaku zistená u i-tej štatistickej jednotky $(i = 1,2, \dots, n)$

n je rozsah súboru, t.j. počet zistených hodnôt v danom súbore

Takto vypočítaný priemer nazývame jednoduchý aritmetický priemer. [5]

Geometrický priemer ($\bar{x}_G$) je n-tá odmocnina zo súčinu jednotlivých hodnôt znaku X: $(x_1, x_2, \dots, x_n)$. Jednoduchý geometrický priemer vypočítame pomocou vzorca [5]:

$$\bar{x}_G = \sqrt[n]{x_1 . x_2 . \dots . x_n} = \sqrt[n]{\prod_{i=1}^{n} x_i} \tag{3.2}$$

3.1.2 Kvantily

Kvantily sú také číselné hodnoty, ktoré rozdeľujú vzostupne usporiadaný štatistický súbor hodnôt na α rovnako početných častí. [5]

Medzi najpoužívanejšie kvantily patria [5] :

- medián, pre $\alpha = 2$
- kvartily, pre $\alpha = 4$
- decily, pre $\alpha = 10$
- percentily, pre $\alpha = 100$

Medián $(\tilde{x}, Me)$ je taká hodnota znaku, ktorá rozdelí súbor vzostupne usporiadaných hodnôt na dve rovnako početné časti. [5]

Kvartily sú také tri hodnoty, ktoré rozdeľujú štatistický súbor na štyri rovnako početné časti, pričom každá obsahuje štvrtinu jednotiek z celkového rozsahu štatistického súboru, t. j. 25 percent štatistických jednotiek. [5]

Dolný kvartil $\left(x_{0,25}, Q_1^4\right)$ oddeľuje štvrtinu jednotiek s najnižšími hodnotami znaku od troch štvrtín štatistických jednotiek s hodnotami vyššími alebo rovnajúcimi sa hodnote dolného kvartilu. Prostredný kvartil $\left(x_{0,50}, \tilde{x}, Me\right)$, je mediánom. Horný kvartil $(x_{0,75}, Q_3^4)$ oddeľuje tri štrvtiny hodnôt nižších alebo rovnajúcich sa a jednu štvrtinu hodnôt vyšších alebo rovnajúcich sa jeho hodnote. [5]

Variačné rozpätie (R) je rozdiel medzi maximálnou a minimálnou hodnotou znaku. Vypočítame ho pomocou vzťahu [5] :

$$R = x_{max} - x_{min} \tag{3.3}$$

Kvantilové rozpätie (R_Q^α) je definované ako rozdiel medzi horným a dolným kvantilom rovnakého druhu. Za dolný kvantil považujeme pri všetkých kvantiloch najmenší z nich, teda prvý (Q_1^α). Horným kvantilom je najväčší z danej skupiny ($Q_{\alpha-1}^\alpha$), teda tretí kvartil, deviaty decil a deväťdesiaty deviaty percentil. [5]

Kvartilové rozpätie je možné vypočítať pomocou vzťahu [5]:

$$R_Q^4 = x_{0,75} - x_{0,25} = Q_3^4 - Q_1^4 \tag{3.4}$$

Kvartilová miera šikmosti S_Q je založená na vzdialenostiach medzi jednotlivými kvartilmi; na jej výpočet stačí poznať ich hodnoty. Počíta sa pomocou vzťahu [5]:

$$S_Q = \frac{(x_{0,75}-x_{0,50})-(x_{0,50}-x_{0,25})}{(x_{0,75}-x_{0,50})+(x_{0,50}-x_{0,25})} \tag{3.5}$$

ktorý je možné upraviť na tvar [5]:

$$S_Q = \frac{x_{0,75}+x_{0,25}-2x_{0,50}}{x_{0,75}-x_{0,25}} \tag{3.6}$$

3.1.3 Časové rady

Časový rad je chronologicky usporiadaná postupnosť porovnateľných kvantitatívnych údajov o skúmanom jave. [5]

Absolútny prírastok (úbytok) (1. diferencia) Δ_t vyjadruje rozdiel medzi dvoma za sebou idúcimi hodnotami časovej premennej. Vyjadruje o koľko sa zvýšila ($\Delta_t > 0$) alebo znížila ($\Delta_t < 0$) hodnota ukazovateľa v období t oproti predchádzajúcemu obdobiu $t-1$. Počíta sa pomocou vzťahu [5]:

$$\Delta_t = y_t - y_{t-1}, \text{pre } t = (2, 3, \dots, T) \tag{3.7}$$

Koeficient rastu k_t je podiel hodnoty ukazovateľa v období t a hodnoty v predchádzajúcom období. Vyjadruje, koľkokrát sa zvýšila ($k_t > 0$) alebo znížila ($k_t < 0$) hodnota ukazovateľa v období t oproti predchádzajúcemu obdobiu $t-1$. Počíta sa pomocou vzťahu [5]:

$$k_t = \frac{y_t}{y_{t-1}}, \text{pre } t = (2, 3, \dots, T) \tag{3.8}$$

Tempo rastu T_t je koeficient rastu vyjadrený v percentách [5]:

$$T_t = k_t \,.\, 100, \text{pre } t = (2, 3, \dots, T) \tag{3.9}$$

Koeficient prírastku k_{Δ_t} je podiel absolútneho prírastku v období t a hodnoty ukazovateľa v období $t-1$. Počíta sa pomocou vzorca [5]:

$$k_{\Delta_t} = \frac{\Delta_t}{y_{t-1}} = \frac{y_t - y_{t-1}}{y_{t-1}} = \frac{y_t}{y_{t-1}} - 1 = k_t - 1, \text{pre } t = (2, 3, \dots, T) \tag{3.10}$$

Tempo prírastku T_{Δ_t} je koeficient prírastku vyjadrený v percentách. Počíta sa pomocou vzťahu [5]:

$$T_{\Delta_t} = k_{\Delta_t}.100 = k_i.100 - 100 = T_i - 100, \text{pre t} = (2, 3, \dots, T) \tag{3.11}$$

Bázický index B_t vyjadruje relatívnu zmenu hodnoty y_t oproti hodnote y_0, ktorú považujeme za bázu (základ) pozorovania. Bázickým obdobím býva väčšinou prvé obdobie v časovom rade, ale môže to byť aj iné obdobie (môže byť aj mimo rozsahu časového radu). Bázický index sa počíta pomocou vzťahu [5]:

$$B_t = \frac{y_t}{y_0}, \text{pre t} = (2, 3, \dots, T) \tag{3.12}$$

Priemerný absolútny prírastok $\bar{\Delta}$ je aritmetickým priemerom absolútnych prírastkov. Vyjadruje, o koľko v priemere vzrástla (klesla) hodnota sledovaného ukazovateľa za jedno časové obdobie. Vypočítame ho pomocou vzťahu [5]:

$$\bar{\Delta} = \frac{\sum_{t=2}^{T} \Delta_t}{T-1} = \frac{(y_2 - y_1) + (y_3 - y_2) + \cdots + (y_T - y_{T-1})}{T-1} = \frac{y_T - y_1}{T-1} \tag{3.13}$$

Priemerný koeficient rastu $\bar{k}$ je geometrickým priemerom jednotlivých koeficientov rastu. Vyjadruje, koľkokrát v priemere vzrástla (klesla) hodnota sledovanej premennej za jedno časové obdobie. Vypočítame ho pomocou vzťahu [5]:

$$\bar{k} = \sqrt[T-1]{k_2.k_3 \dots .k_T} = \sqrt[T-1]{\frac{y_2}{y_1}.\frac{y_3}{y_2}.\dots.\frac{y_T}{y_{T-1}}} = \sqrt[T-1]{\frac{y_T}{y_1}} \tag{3.14}$$

Priemerné tempo rastu $\bar{T}$ je priemerný koeficient rastu vyjadrený v percentách. Vypočítame ho pomocou vzťahu [5]:

$$\bar{T} = \bar{k}.100 \tag{3.15}$$

Priemerné tempo prírastku $\bar{T}_{\Delta}$ dostaneme, ak od priemerného tempa rastu odpočítame 100. Vyjadruje, o koľko percent v priemere za jedno časové obdobie rástli (klesali) hodnoty časového radu v sledovanom časovom rade [5]:

$$\bar{T}_{\Delta} = \bar{T} - 100 \tag{3.16}$$

3.2 Typy grafov

Bodový graf sa používa na zobrazenie výsledkov triedenia podľa jedného číselného znaku v súbore štatistických jednotiek. Body tohto grafu vznikajú ako priesečníky príslušných hodnôt znaku, zaznamenaných na osi *x* a ich absolútnych, resp. relatívnych početností na osi *y*. Bodovým grafom možno znázorniť aj časový rad: na osi *x* je časová premenná a na osi *y* hodnoty ukazovateľa v jednotlivých časových obdobiach. [5]

Spojnicový graf je lomená čiara, ktorú zostrojíme pospájaním niektorých bodov v bodovom grafe. Ak bodový graf znázorňuje rozdelenie početnosti číselného znaku (na osi *x* sú hodnoty znaku a na hodnote *y* absolútne, resp. relatívne početnosti), vzniknutý spojnicový graf voláme polygón rozdelenia početností. [5]

Kruhový (koláčový) graf umožňuje grafické znázornenie diskrétneho číselného alebo slovného znaku. Obsah výseku (v porovnaní s obsahom celého kruhu) vyjadruje podiel štatistických jednotiek s danou obmenou znaku (alebo s číselnou hodnotou) . [5]

Histogram je stĺpcový diagram, ktorý je vhodný na grafické znázornenie rozdelenia početností kvalitatívneho znaku, ako aj intervalového rozdelenia početností. Tvoria ho obdĺžniky, ktorých výška je priamoúmerná početnostiam príslušnej obmeny znaku, prípadne príslušných intervalov. [5]

Stĺpcový graf alebo stĺpcový diagram je diagram, ktorý znázorňuje zloženie sledovaného súboru pomocou obdĺžnikových pruhov, ktorých dĺžka proporcionálne zodpovedá veľkosti hodnôt, ktoré znázorňujú. Pruhy môžu byť nakreslené zvisle aj vodorovne. [6]

Škatuľkový graf (box and whisker plot) má tvar obdĺžnika (krabice) umiestneného nad súradnicovou osou *x*. Ľavá (pravá) strana je umiestnená nad dolným (horným) kvartilom. Zvislá čiara predeľujúca obdĺžnik je nad hodnotou mediánu a krížik nad aritmetickým priemerom. Z obdĺžnika po ľavej aj pravej strane vychádzajú úsečky, ktorých koncový bod je nad minimálnou, resp. maximálnou hodnotou. [5]

Tepelná mapa je technika vizualizácie údajov, ktorá zobrazuje veľkosť javu ako farbu v dvoch rozmeroch. Zmena farby môže byť odtieňom alebo intenzitou, čo dáva čitateľovi jasné vizuálne podnety o tom, ako je jav zoskupený alebo sa mení v priestore. [7]

Kartogramy sa využívajú najmä pri štrukturálnych analýzach z priestorového hľadiska. Vyjadrujú rozmiestnenie sledovaného ukazovateľa v priestore. Sú to mapy rozdelené podľa určitého kritéria na oblasti, ktoré rozlišujeme vzhľadom na intenzitu sledovaného javu farebne alebo pomocou textúry. (napr. šrafovanie) [5]

4 Výsledky práce

V tejto kapitole sú uvedené zdroje všetkých údajov použitých v tejto práci, tieto sa nachádzajú v podkapitolách nižšie (podkapitoly 4.2.1, 4.2.2, 4.2.3 a 4.8.1) a v ďalších podkapitolách sú postupne uvedené príklady pre vybrané typy grafov s použitím týchto dát, pričom sú k tomu použité viaceré knižnice programovacieho jazyka Python, ktoré sú určené k spracovaniu a vizualizácii dát. Nachádzajú sa tu základné a takisto aj pokročilejšie grafické zobrazenia vytvorené funkciami knižníc *seaborn*, *matplotlib* a *geopandas* pri použití knižnice *pandas* na prvotný výber a spracovanie dát.

4.1 Import knižníc a načítanie dát

Na nasledujúcom obrázku (Obrázok 2) je zobrazený iniciálny zdrojový kód, ktorý obsahuje import knižníc, ktoré sú použité ďalej v zdrojovom kóde, konkrétne sú to knižnice *numpy*, *pandas*, *seaborn*, *matplotlib* a jeho rozhrania *matplotlib.pyplot* a knižnice *geopandas* použitej neskôr pri vykresľovaní mapy. Ďalšie časti kódu sa venujú načítaniu dát použitých pri vykresľovaní grafov, pričom tieto dáta sú bližšie popísané v podkapitolách nižšie (4.2.1, 4.2.2, 4.2.3 a 4.8.1).

Po úvodnom zdrojovom kóde pre naimportovanie knižníc, ktoré sú neskôr použité v ďalších častiach zdrojového kódu, nasleduje časť zdrojového kódu, ktorá načítava údaje o krajinách zo súboru who_life_exp.csv pomocou knižnice *pandas* do objektu typu *DataFrame* a takisto vytvorí ďalší objekt typu *DataFrame*, ktorý obsahuje dáta o jednotlivých krajinách iba z roku 2016, keďže vo viacerých zobrazeniach nižšie budú použité dáta výlučne pre daný rok.

Ďalej nasleduje načítanie slovenských názvov krajín a ich kódov vo formáte

ISO3166-1-Alpha-3 zo súboru iso_a3_sk.csv pomocou knižnice *pandas* najskôr do objektu typu *DataFrame*, pričom z hodnôt, ktoré sa v ňom nachádzajú je ďalej vytvorený slovník, kde kľúče sú hodnoty kódov a jeho hodnoty sú slovenské názvy jednotlivých krajín načítaných z daného súboru, čo umožňuje pripojiť slovenské názvy týchto krajín k už existujúcim dátam, čo je

potrebné z dôvodu, že pôvodný dataset obsahuje názvy použitých krajín výhradne v anglickom jazyku.

Ďalšia časť kódu načítava geografické údaje o krajinách sveta zo súboru countries.geojson, ktorý je vo formáte GeoJSON, pomocou knižnice *geopandas* do objektu typu GeoDataFrame a pripája k týmto údajom informácie zo skôr načítaného datasetu s údajmi z roku 2016 pre umožnenie zobrazenia týchto údajov na mape.

Záverečná časť tohto zdrojového kódu načítava do pamäte súbor city_temperature.csv obsahujúci dáta o priemerných mesačných teplotách vo vybraných mestách pomocou knižnice *pandas*.

Obrázok 2 **Zdrojový kód pre importovanie knižníc a načítanie vstupných dát**

```
In [1]: import numpy as np
        import pandas as pd
        import seaborn as sns
        import matplotlib.pyplot as plt
        import geopandas as gpd

        #Načítanie vybraných informácií o krajinách a vytiahnutie údajov špecificky z roku 2016
        who_df = pd.read_csv(r"who_life_exp.csv", usecols = ["country", "country_code", "region", "year",
                                                            "life_expect","alcohol", "basic_water", "une_pop"])
        who_2016_df = who_df[who_df["year"] == 2016]

        # Načítanie slovenských názvov krajín a ich uloženie do slovníka
        country_sk_df = pd.read_csv(r"iso_a3_sk.csv", usecols = ["alpha3", "country_name"])
        country_sk_df.set_index("alpha3", inplace = True)
        country_sk_dict = dict(country_sk_df["country_name"])

        # Načítanie geografických dát a pripojenie údajov o krajinách z roku 2016
        map_df = gpd.read_file(r"countries.geojson", usecols = ["ISO_A3", "Geometry"])
        map_data_df = map_df.merge(right = who_2016_df, left_on = "ISO_A3",
                                   right_on = "country_code", how = "left", indicator = True)

        # Načítanie dát o počasí v mestách
        weather_df = pd.read_csv(r"city_temperature.csv", usecols = ['City', 'Month', 'Day', 'Year', 'AvgTemperature'])
```

Zdroj: Vlastné spracovanie

4.2 Použité dáta

V tejto podkapitole sú uvedené dáta použité pri grafických zobrazeniach, ktoré spoločne so zdrojovým kódom použitým pre ich vykreslenie sa nachádzajú v ďalších podkapitolách tejto kapitoly.

4.2.1 Štatistické údaje o krajinách

Ako vstupné dáta pre analýzu údajov o jednotlivých krajinách slúžia dáta umiestnené na webovej adrese https://www.kaggle.com/mmattson/who-national-life-expectancy, odkiaľ bol stiahnutý súbor s názvom who_life_exp.csv. Tento dataset obsahuje informácie o 183 krajinách sveta pochádzajúce z rokov 2000 až 2016. Tento dataset bol vytvorený s použitím informácií poskytnutých organizáciami GHO (Globálne observatórium zdravia) a UNESCO (Organizácia OSN pre vzdelávanie, vedu a kultúru). Z týchto sú ďalej použité stĺpce zobrazené v tabuľke nižšie (Tabuľka 1). [19]

Tabuľka 1 **Zoznam použitých premenných zo súboru who_life_exp.csv**

Názov premennej	**Popis**
country	názov krajiny v anglickom jazyku
country_code	kód krajiny vo formáte ISO3166-1-Alpha-3
region	región
year	rok
life_expect	priemerná dĺžka života
alcohol	spotreba alkoholu na obyvateľa vo veku 15+ v litroch
basic_water	percentuálny podiel obyvateľov s prístupom k pitnej vode
une_pop	populácia v tisícoch

Zdroj: Vlastné spracovanie podľa údajov z www.kaggle.com

4.2.2 Priemerná denná teplota

Pre informácie o priemernej dennej teplote vo vybraných mestách je v práci použitý dataset, ktorý sa nachádza na webovej stránke www.kaggle.com, konkrétne na adrese

https://www.kaggle.com/datasets/sudalairajkumar/daily-temperature-of-major-cities, kde sa nachádza súbor city_temperature.csv, ktorý sa nachádza v archíve city_temperature.zip a

obsahuje daný dataset. Tento dataset obsahuje informácie o 321 mestách sveta a ich priemernej dennej teplote pre dátumy od 1. januára 1995 do 13. mája 2020. [19]

Tabuľka 2 **Zoznam použitých premenných zo súboru city_temperature.csv**

Názov premennej	**Popis**
City	názov mesta
Month	Mesiac
Day	Deň
Year	Rok
AvgTemperature	Priemerná teplota v danom mesiaci

Zdroj: Vlastné spracovanie podľa údajov z www.kaggle.com

4.2.3 Názvy krajín v slovenskom jazyku

Názvy krajín v slovenskom jazyku podľa kódu ISO3166-1-Alpha-3 boli použité na základe informácií nachádzajúcich sa na adrese https://sk.wikipedia.org/wiki/ISO_3166-1. Tento kód spoločne so slovenským názvom krajín boli uložené do súboru iso_a3_sk.csv pod zmenenými názvami. Stĺpce výsledného súboru sú uvedené v nasledujúcej tabuľke (Tabuľka 3). [25]

Tabuľka 3 **Zoznam premenných súboru iso_a3_sk.csv**

Názov premennej	**Popis**
alpha3	kód krajiny vo formáte ISO3166-1-Alpha-3
country_name	slovenský názov krajiny

Zdroj: Vlastné spracovanie podľa údajov z sk.wikipedia.org

4.3 Spojnicový graf

Na obrázku nižšie (Obrázok 3) je zobrazený zdrojový kód pre vykreslenie grafu, ktorý zobrazuje priemernú ročnú spotrebu čistého alkoholu v litroch na osobu pre osoby vo veku 15 a viac rokov v krajinách Vyšehradskej štvorky v rokoch 2000 až 2016. Prvá časť zdrojového kódu vyťahuje údaje o spotrebe alkoholu do samostatného objektu typu *DataFrame* s názvom *alcohol_df*. Ďalej sa tu pre jednotlivé krajiny Vyšehradskej štvorky vyberajú hodnoty pre jednotlivé krajiny a postupne sa vykresľujú, pričom na vykreslenie grafu je použitá funkcia *plot*, ktorá sa nachádza v knižnici *matplotlib*. Ďalej sa tu nachádza zobrazenie popisov jednotlivých osí pomocou funkcií *xlabel* a *ylabel* z knižnice *matplotlib* s použitím modrého písma. Na konci je vytvorenie legendy pomocou funkcie *legend* nachádzajúcej sa v knižnici *matplotlib*, ktorá je zobrazená v pravom dolnom rohu.

Obrázok 3 **Zdrojový kód pre vytvorenie spojnicového grafu spotreby alkoholu V4**

```
In [2]: #Vytvorenie možnosti zobrazenia viacerých grafov súčasne
        fig, ax = plt.subplots(figsize = (9, 6))

        alcohol_df = who_df[["year", "alcohol"]]

        #Vykreslenie spotreby alkoholu na Slovensku
        alcohol_svk_df = who_df[who_df["country_code"] == "SVK"]
        ax.plot(alcohol_svk_df["year"], alcohol_svk_df["alcohol"], color = "blue", label = "Slovensko")

        #Vykreslenie spotreby alkoholu v Česku
        alcohol_cze_df = who_df[who_df["country_code"] == "CZE"]
        ax.plot(alcohol_cze_df["year"], alcohol_cze_df["alcohol"], color = "purple", label = "Česko")

        #Vykreslenie spotreby alkoholu v Poľsku
        alcohol_pol_df = who_df[who_df["country_code"] == "POL"]
        ax.plot(alcohol_pol_df["year"], alcohol_pol_df["alcohol"], color = "red", label = "Poľsko")

        #Vykreslenie spotreby alkoholu v Maďarsku
        alcohol_hun_df = who_df[who_df["country_code"] == "HUN"]
        ax.plot(alcohol_hun_df["year"], alcohol_hun_df["alcohol"], color = "green", label = "Maďarsko")

        #Označenie osí x a y
        plt.xlabel("Rok", weight = "bold", color = "blue")
        plt.ylabel("Ročná spotreba alkoholu na osobu (l)", weight = "bold", color = "blue")

        #Zobrazenie legendy
        plt.legend(loc = "lower right")

        plt.show()
```

Zdroj: Vlastné spracovanie

Ako je vidieť na grafe (Graf 1), v roku 2000 malo z krajín Vyšehradskej štvorky priemernú ročnú spotrebu čistého alkoholu v litroch na osobu pre osoby vo veku 15 a viac rokov Česko, o niečo menej malo Maďarsko, ešte menej malo Slovensko a Poľsko malo z týchto krajín jednoznačne najnižšiu spotrebu. V ďalších rokoch sa Česko držalo medzi týmito krajinami celé sledované obdobie na prvom mieste, pričom jeho spotreba mala väčšinu daného obdobia mierne klesajúcu tendenciu.

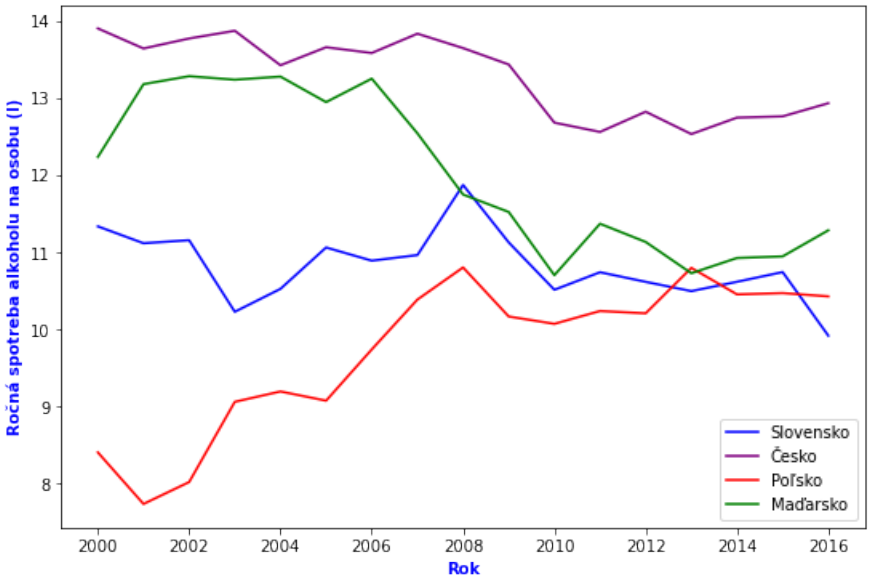
V Maďarsku spotreba stúpala v rokoch 2000 až 2006, pričom v rokoch 2004 a 2006 bola spotreba veľmi blízko spotrebe v Česku. V ďalšom období v rokoch 2006 až 2016 spotreba väčšinou klesala, pričom okrem rokov 2008 a 2013 bolo Maďarsko v tejto spotrebe celý čas na druhom mieste zo sledovaných krajín.

Na Slovensku nie je na grafe z dlhodobého hľadiska viditeľný jednoznačný stúpajúci alebo klesajúci trend, mierny jednorazový nárast bol zaznamenaný v roku 2008, kedy bola spotreba mierne vyššia v porovnaní s Maďarskom, pokles je viditeľný v rokoch 2003 a 2016.

V Poľsku bol zaznamenaný mierny pokles medzi rokmi 2000 a 2001, potom v období 2001 až 2008 bolo zaznamenané výrazné stúpanie spotreby alkoholu. V ďalšom období v rokoch 2008 až 2016 nie je viditeľný jednoznačný trend vo vývoji spotreby alkoholu. V roku 2013 bol zaznamenaný jednorazový nárast, v dôsledku mierneho poklesu v Maďarsku a na Slovensku sa Poľsko nachádzalo v danom roku na druhom mieste zo sledovaných krajín. Na konci sledovaného obdobia v roku 2016 sa Poľsko dostalo na tretie miesto zo sledovaných krajín pred Slovensko z dôvodu poklesu spotreby na Slovensku oproti roku 2015.

Ďalší vývoj spotreby alkoholu v týchto krajinách po roku 2016 nie je zo sledovaných dát možné vyčítať.

Graf 1 **Spojnicový graf - ročná spotreba alkoholu na osobu v krajinách V4**



Zdroj: Vlastné spracovanie podľa údajov z www.kaggle.com

Nasledujúce tabuľky (Tabuľka 4, Tabuľka 5, Tabuľka 6, Tabuľka 7) uvádzajú základné charakteristiky časových radov postupne pre Slovensko, Česko, Poľsko a Maďarsko, pričom tieto časové rady sú tvorené hodnotami priemernej spotreby alkoholu na osobu pre osoby vo veku 15 a viac rokov v litroch v daných krajinách pre roky 2000 až 2016 a ktorých hodnoty sú znázornené aj na predchádzajúcom grafe (Graf 1).

Tabuľka 4 **Časový rad spotreby alkoholu na obyvateľa v litroch na Slovensku**

Rok	Spotreba alkoholu	t	Δ_t	k_t	T_t	k_{Δ_t}	T_{Δ_t}	B_t
2000	11.33439	1						1.00000
2001	11.11560	2	-0.21879	0.98070	98.06968	-0.01930	-1.93032	0.98070
2002	11.15501	3	0.03941	1.00355	100.35455	0.00355	0.35455	0.98417
2003	10.22684	4	-0.92817	0.91679	91.67934	-0.08321	-8.32066	0.90228
2004	10.52383	5	0.29699	1.02904	102.90403	0.02904	2.90403	0.92849
2005	11.06067	6	0.53684	1.05101	105.10118	0.05101	5.10118	0.97585
2006	10.88943	7	-0.17124	0.98452	98.45181	-0.01548	-1.54819	0.96074
2007	10.96214	8	0.07271	1.00668	100.66771	0.00668	0.66771	0.96716
2008	11.87122	9	0.90908	1.08293	108.29291	0.08293	8.29291	1.04736
2009	11.12618	10	-0.74504	0.93724	93.72398	-0.06276	-6.27602	0.98163
2010	10.51203	11	-0.61415	0.94480	94.48014	-0.05520	-5.51986	0.92745
2011	10.73959	12	0.22756	1.02165	102.16476	0.02165	2.16476	0.94752
2012	10.61502	13	-0.12457	0.98840	98.84009	-0.01160	-1.15991	0.93653
2013	10.49467	14	-0.12035	0.98866	98.86623	-0.01134	-1.13377	0.92591
2014	10.61569	15	0.12102	1.01153	101.15316	0.01153	1.15316	0.93659
2015	10.74172	16	0.12603	1.01187	101.18720	0.01187	1.18720	0.94771
2016	9.91722	17	-0.82450	0.92324	92.32432	-0.07676	-7.67568	0.87497

Zdroj: Vlastné spracovanie podľa údajov z www.kaggle.com

Z predchádzajúcej tabuľky (Tabuľka 4) pre uvedené hodnoty charakteristík časového radu vyplývajú pre rok 2002 nasledujúce skutočnosti : V roku 2002 stúpla spotreba čistého alkoholu na osobu pre osoby vo veku 15 a viac rokov na Slovensku v porovnaní s rokom 2001 o 0.03941 litra. V roku 2002 došlo v porovnaní s rokom 2001 k nárastu spotreby čistého alkoholu na osobu pre osoby vo veku 15 a viac rokov na Slovensku o 0.35455 percenta. V roku 2002 poklesla spotreba čistého alkoholu na osobu pre osoby vo veku 15 a viac rokov na Slovensku o 1.583 percenta v porovnaní s rokom 2000.

Tabuľka 5 **Časový rad spotreby alkoholu na obyvateľa v litroch v Česku**

Rok	Spotreba alkoholu	t	Δ_t	k_t	T_t	k_{Δ_t}	T_{Δ_t}	B_t
2000	13.90077	1						1.00000
2001	13.64064	2	-0.26013	0.98129	98.12866	-0.01871	-1.87134	0.98129
2002	13.77038	3	0.12974	1.00951	100.95113	0.00951	0.95113	0.99062
2003	13.87036	4	0.09998	1.00726	100.72605	0.00726	0.72605	0.99781
2004	13.42303	5	-0.44733	0.96775	96.77492	-0.03225	-3.22508	0.96563
2005	13.65681	6	0.23378	1.01742	101.74163	0.01742	1.74163	0.98245
2006	13.58224	7	-0.07457	0.99454	99.45397	-0.00546	-0.54603	0.97709
2007	13.83238	8	0.25014	1.01842	101.84167	0.01842	1.84167	0.99508
2008	13.64567	9	-0.18671	0.98650	98.65020	-0.01350	-1.34980	0.98165
2009	13.43307	10	-0.21260	0.98442	98.44200	-0.01558	-1.55800	0.96635
2010	12.67902	11	-0.75405	0.94387	94.38661	-0.05613	-5.61339	0.91211
2011	12.55871	12	-0.12031	0.99051	99.05111	-0.00949	-0.94889	0.90345
2012	12.82090	13	0.26219	1.02088	102.08771	0.02088	2.08771	0.92232
2013	12.53090	14	-0.29000	0.97738	97.73807	-0.02262	-2.26193	0.90145
2014	12.74424	15	0.21334	1.01703	101.70251	0.01703	1.70251	0.91680
2015	12.76054	16	0.01630	1.00128	100.12790	0.00128	0.12790	0.91797
2016	12.93089	17	0.17035	1.01335	101.33497	0.01335	1.33497	0.93023

Zdroj: Vlastné spracovanie podľa údajov z www.kaggle.com

Z predchádzajúcej tabuľky (Tabuľka 5) pre uvedené hodnoty charakteristík časového radu vyplývajú pre rok 2002 nasledujúce skutočnosti : V roku 2002 stúpla spotreba čistého alkoholu na osobu pre osoby vo veku 15 a viac rokov v Česku v porovnaní s rokom 2001 o 0.12974 litra. V roku 2002 došlo v porovnaní s rokom 2001 k nárastu spotreby čistého alkoholu na osobu pre osoby vo veku 15 a viac rokov v Česku o 0.95113 percenta. V roku 2002 poklesla spotreba čistého alkoholu na osobu pre osoby vo veku 15 a viac rokov v Česku o 0.938 percenta v porovnaní s rokom 2000.

Tabuľka 6 **Časový rad spotreby alkoholu na obyvateľa v litroch v Poľsku**

Rok	Spotreba alkoholu	t	Δ_t	k_t	T_t	k_{Δ_t}	T_{Δ_t}	B_t
2000	8.40411	1						1.00000
2001	7.73515	2	-0.66896	0.92040	92.04009	-0.07960	-7.95991	0.92040
2002	8.01907	3	0.28392	1.03671	103.67052	0.03671	3.67052	0.95418
2003	9.05941	4	1.04034	1.12973	112.97332	0.12973	12.97332	1.07797
2004	9.19369	5	0.13428	1.01482	101.48222	0.01482	1.48222	1.09395
2005	9.07429	6	-0.11940	0.98701	98.70128	-0.01299	-1.29872	1.07974
2006	9.73869	7	0.66440	1.07322	107.32178	0.07322	7.32178	1.15880
2007	10.38513	8	0.64644	1.06638	106.63785	0.06638	6.63785	1.23572
2008	10.80180	9	0.41667	1.04012	104.01218	0.04012	4.01218	1.28530
2009	10.16544	10	-0.63636	0.94109	94.10876	-0.05891	-5.89124	1.20958
2010	10.07045	11	-0.09499	0.99066	99.06556	-0.00934	-0.93444	1.19828
2011	10.23680	12	0.16635	1.01652	101.65186	0.01652	1.65186	1.21807
2012	10.20692	13	-0.02988	0.99708	99.70811	-0.00292	-0.29189	1.21452
2013	10.79606	14	0.58914	1.05772	105.77197	0.05772	5.77197	1.28462
2014	10.45218	15	-0.34388	0.96815	96.81476	-0.03185	-3.18524	1.24370
2015	10.46922	16	0.01704	1.00163	100.16303	0.00163	0.16303	1.24573
2016	10.42659	17	-0.04263	0.99593	99.59281	-0.00407	-0.40719	1.24065

Zdroj: Vlastné spracovanie podľa údajov z www.kaggle.com

Z predchádzajúcej tabuľky (Tabuľka 6) pre uvedené hodnoty charakteristík časového radu vyplývajú pre rok 2002 nasledujúce skutočnosti : V roku 2002 stúpla spotreba čistého alkoholu na osobu pre osoby vo veku 15 a viac rokov v Poľsku v porovnaní s rokom 2001 o 0.28392 litra. V roku 2002 došlo v porovnaní s rokom 2001 k nárastu spotreby čistého alkoholu na osobu pre osoby vo veku 15 a viac rokov v Poľsku o 3.67052 percenta. V roku 2002 poklesla spotreba čistého alkoholu na osobu pre osoby vo veku 15 a viac rokov v Poľsku o 4.582 percenta v porovnaní s rokom 2000.

Tabuľka 7 **Časový rad spotreby alkoholu na obyvateľa v litroch v Maďarsku**

Rok	Spotreba alkoholu	t	Δ_t	k_t	T_t	k_{Δ_t}	T_{Δ_t}	B_t
2000	12.23341	1						1.00000
2001	13.17809	2	0.94468	1.07722	107.72213	0.07722	7.72213	1.07722
2002	13.28141	3	0.10332	1.00784	100.78403	0.00784	0.78403	1.08567
2003	13.23592	4	-0.04549	0.99657	99.65749	-0.00343	-0.34251	1.08195
2004	13.27615	5	0.04023	1.00304	100.30395	0.00304	0.30395	1.08524
2005	12.94530	6	-0.33085	0.97508	97.50794	-0.02492	-2.49206	1.05819
2006	13.24966	7	0.30436	1.02351	102.35112	0.02351	2.35112	1.08307
2007	12.54053	8	-0.70913	0.94648	94.64794	-0.05352	-5.35206	1.02511
2008	11.74712	9	-0.79341	0.93673	93.67323	-0.06327	-6.32677	0.96025
2009	11.52080	10	-0.22632	0.98073	98.07340	-0.01927	-1.92660	0.94175
2010	10.70125	11	-0.81955	0.92886	92.88634	-0.07114	-7.11366	0.87476
2011	11.36837	12	0.66712	1.06234	106.23404	0.06234	6.23404	0.92929
2012	11.13217	13	-0.23620	0.97922	97.92231	-0.02078	-2.07769	0.90998
2013	10.72647	14	-0.40570	0.96356	96.35561	-0.03644	-3.64439	0.87682
2014	10.92435	15	0.19788	1.01845	101.84478	0.01845	1.84478	0.89299
2015	10.94509	16	0.02074	1.00190	100.18985	0.00190	0.18985	0.89469
2016	11.28316	17	0.33807	1.03089	103.08878	0.03089	3.08878	0.92232

Zdroj: Vlastné spracovanie podľa údajov z www.kaggle.com

Z predchádzajúcej tabuľky (Tabuľka 7) pre uvedené hodnoty charakteristík časového radu vyplývajú pre rok 2002 nasledujúce skutočnosti: V roku 2002 stúpla spotreba čistého alkoholu na osobu pre osoby vo veku 15 a viac rokov v Maďarsku v porovnaní s rokom 2001 o 0.10332 litra. V roku 2002 došlo v porovnaní s rokom 2001 k nárastu spotreby čistého alkoholu na osobu pre osoby vo veku 15 a viac rokov v Maďarsku o 0.78403 percenta. V roku 2002 stúpla spotreba čistého alkoholu na osobu pre osoby vo veku 15 a viac rokov v Maďarsku o 8.567 percenta v porovnaní s rokom 2000.

Nasledujúca tabuľka (Tabuľka 8) zobrazuje priemery základných charakteristík časových radov, ktoré sú tvorené hodnotami priemernej spotreby alkoholu na osobu pre osoby vo veku 15 a viac rokov v litroch v krajinách Vyšehradskej štvorky pre roky 2000 až 2016 uvedených v tabuľkách vyššie (Tabuľka 4, Tabuľka 5, Tabuľka 6 a Tabuľka 7).

Tabuľka 8 **Priemery charakteristík časových radov spotreby alkoholu**

Krajina	$\bar{\Delta}$	$\bar{k}$	$\bar{T}$	$\bar{T}_{\Delta}$
Slovensko	-0.08857	0.99169	99.16867	-0.83133
Česko	-0.06062	0.99549	99.54899	-0.45101
Poľsko	0.12641	1.01357	101.35686	1.35686
Maďarsko	-0.05939	0.99495	99.49590	-0.50410

Zdroj: Vlastné spracovanie podľa údajov z www.kaggle.com

Z predchádzajúcej tabuľky (Tabuľka 8) pre uvedené priemery základných charakteristík časového radu pre Slovensko vyplývajú nasledujúce skutočnosti: Počas sledovaného obdobia klesala priemerná ročná spotreba čistého alkoholu na osobu pre osoby vo veku 15 a viac rokov na Slovensku v priemere o 0.08857 litra ročne. Počas sledovaného obdobia klesala priemerná ročná spotreba čistého alkoholu na osobu pre osoby vo veku 15 a viac rokov na Slovensku v priemere o 0.83133 percenta ročne.

4.4 Stĺpcový graf

Na obrázku nižšie (Obrázok 4) je zobrazený zdrojový kód pre vykreslenie stĺpcového grafu, ktorý zobrazuje priemernú ročnú spotrebu čistého alkoholu v litroch na osobu pre osoby vo veku 15 a viac rokov v roku 2016 v 10 krajinách, ktoré majú zo skúmaných krajín túto

spotrebu najvyššiu, pričom daná spotreba sa tu zaokrúhľuje na 2 desatinné miesta. Prvá časť zdrojového kódu sa venuje zaokrúhleniu hodnôt spotreby alkoholu v roku 2016 na 2 desatinné miesta a výberu 10 krajín s najväčšou spotrebou, pričom pre zoradenie týchto hodnôt je tu použitá funkcia *sort_values* nachádzajúca sa v objekte *DataFrame*, ktorý sa nachádza v knižnici *pandas*. Ďalšia časť zdrojového kódu pridáva k tomuto objektu nový stĺpec obsahujúci názvy krajín v slovenskom jazyku, ktoré sú pridané s použitím slovníka *country_sk_dict*, ktorý bol vytvorený v iniciálnom kóde (Obrázok 2) a tento stĺpec je ďalej nastavený ako index v tomto objekte.

Ďalej sa tu nachádza vykreslenie samotného grafu pomocou funkcie barplot, ktorá sa nachádza v knižnici *seaborn*, pričom pomocou parametra *orientation* je tu nastavená horizontálna orientácia grafu. Ďalej sa v kóde nachádza označenie jednotlivých osí pomocou funkcií *xlabel* a *ylabel*, ktoré sa nachádzajú v knižnici *matplotlib*. Obe tieto označenia sú vykreslené boldom modrou farbou písma. Os *x* obsahuje hodnoty množstva alkoholu v litroch. Na osi *y* sú zobrazené názvy jednotlivých krajín. Ďalej sa tu nachádza kód zabezpečujúci v cykle vypísanie hodnôt spotreby alkoholu pre jednotlivé krajiny pomocou funkcie *text* z knižnice *matplotlib*.

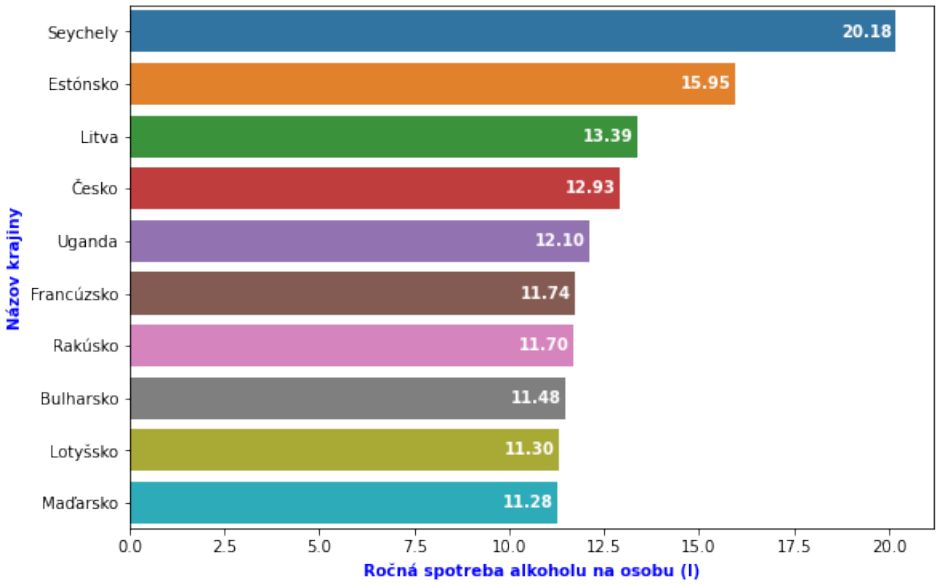
Obrázok 4 **Kód pre vytvorenie stĺpcového grafu spotreby alkoholu (2016)**

```
#Výber 10 krajín s najväčšou spotrebou alkoholu v roku 2016
alcohol_2016_df = who_2016_df[["country", "country_code", "alcohol"]].copy()
alcohol_2016_df["alcohol"] = alcohol_2016_df["alcohol"].apply(round, args = (2,))
alcohol_top_10_df = alcohol_2016_df.sort_values(by = "alcohol", ascending = False)[:10]

#Pridanie slovenských názvov krajín
alcohol_top_10_df["country_sk"] = alcohol_top_10_df["country_code"].map(country_sk_dict)
alcohol_top_10_df.set_index("country_code", inplace = True)

plt.figure(figsize = (9, 6))

#Vykreslenie stĺpcového grafu
sns.barplot(data = alcohol_top_10_df, x = "alcohol", y = "country_sk", orientation = "horizontal")

#Označenie osí x a y
plt.xlabel("Ročná spotreba alkoholu na osobu (l)", weight = "bold", color = "blue")
plt.ylabel("Názov krajiny", weight = "bold", color = "blue")

#Vypísanie číselných hodnôt do grafu
for position, country in enumerate(alcohol_top_10_df.index):
    value = alcohol_top_10_df["alcohol"][country]
    text_x = value - 0.1
    text_y = position
    text = format(value, '.2f')
    plt.text(text_x, text_y, text, va = "center", ha = "right", color = "white", weight = "bold")

plt.show()
```

Zdroj: Vlastné spracovanie

Ako je vidieť na grafe (Graf 2), najvyššia ročná spotreba čistého alkoholu na osobu pre osoby vo veku 15 a viac rokov bola v roku 2016 spomedzi všetkých 183 skúmaných krajín na Seycheloch, kde prestavovala 20.18 litra na osobu. Na druhom mieste bolo Estónsko so spotrebou 15.95 litra na osobu, na treťom mieste Litva so spotrebou 13.39 litra na osobu, na štvrtom mieste bolo Česko so spotrebou 12.93 litra na osobu, na piatom mieste bola Uganda so spotrebou 12.10 litra na osobu, na šiestom mieste Francúzsko so spotrebou 11.74 litra na osobu, na siedmom mieste Rakúsko so spotrebou 11.70 litra na osobu, na ôsmom mieste Bulharsko so spotrebou 11.48 litra na osobu, na deviatom mieste bolo Lotyšsko so spotrebou 11.30 litra na osobu a na desiatom mieste bolo Maďarsko, kde bola spotreba 11.28 litra na osobu. Slovensko sa nachádzalo na dvadsiatom mieste so spotrebou 9.92 litra na osobu. Medzi krajinami s vysokou spotrebou alkoholu prevažujú európske krajiny, pričom z prvých 10 krajín zobrazených na grafe je 8 európskych krajín. Veľká spotreba bola v pobaltských krajinách, ktoré sa všetky nachádzali v prvej desiatke, pričom Estónsko a Litva boli v tomto poradí na druhom a treťom mieste.

Graf 2 **Stĺpcový graf – Krajiny s najväčšou spotrebou alkoholu na osobu (2016)**



Zdroj: Vlastné spracovanie podľa údajov z www.kaggle.com a sk.wikipedia.org

4.5 Kruhový graf

Na nasledujúcom obrázku (Obrázok 5) je zobrazený zdrojový kód pre vytvorenie kruhového grafu znázorňujúceho percentuálne rozloženie obyvateľstva medzi krajinami Vyšehradskej štvorky v roku 2016. Prvá časť zdrojového kódu sa venuje vytvoreniu slovníka, ktorý obsahuje ako kľúče kódy krajín Vyšehradskej štvorky a jeho hodnoty sú farby, ktorými majú byť tieto krajiny na danom kruhovom grafe zobrazené. Ďalej sa tu vyťahujú údaje z roku 2016 o krajinách Vyšehradskej štvorky, zoradia sa podľa počtu obyvateľov a pridá sa k nim pomocou vyššie vytvoreného slovníka informácia o farbách, ktoré budú pre jednotlivé krajiny použité. Ďalej sa v kóde nachádza volanie funkcie *pie* z knižnice *matplotlib*, v ktorom sa nastavujú farby jednotlivých výsekov a zaokrúhlenie zobrazených hodnôt. Kruhové výseky zobrazujú číselný údaj o percentuálnom podiele obyvateľov danej krajiny v pomere k celkovému počtu obyvateľov všetkých štyroch krajín Vyšehradskej štvorky, pričom daná hodnota je zaokrúhlená na 1 desatinné miesto. Ďalšia časť zdrojového kódu pridáva legendu a umiestňuje ju do pravého horného rohu.

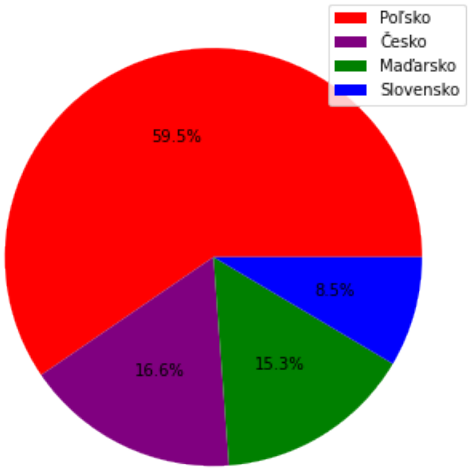
Obrázok 5 **Kód pre vytvorenie kruhového grafu populácie V4 (2016)**

```
In [4]: #Zoznam krajín V4 spolu s ich farebným označením
visegrad_dict = {"SVK" : "blue",
                 "CZE" : "purple",
                 "POL" : "red",
                 "HUN" : "green"}

#Výber údajov o krajinách V4 z roku 2016, zoradenie podľa populácie a pridanie informácie o farbe
pop_df = who_2016_df[who_2016_df["country_code"].isin(visegrad_dict.keys())].copy()
pop_df["country_sk"] = pop_df["country_code"].map(country_sk_dict)
pop_sorted_df = pop_df.sort_values(by = "une_pop", ascending = False)
pop_sorted_df["country_color"] = pop_sorted_df["country_code"].map(visegrad_dict)

plt.figure(figsize = (9, 6))

#Vykreslenie kruhového grafu
plt.pie(x = pop_sorted_df["une_pop"], colors = pop_sorted_df["country_color"], autopct = "%1.1f%%")

#Zobrazenie legendy
plt.legend(labels = pop_sorted_df["country_sk"], loc = "upper right")

plt.show()
```

Zdroj: Vlastné spracovanie

Na grafe (Graf 3) je vidieť, že v roku 2016 tvorili obyvatelia Poľska 59.5 percenta zo všetkých obyvateľov Vyšehradskej štvorky, obyvatelia Česka tvorili 16.6 percenta, obyvatelia Maďarska 15.3 percenta a obyvatelia Slovenska 8.5 percenta.

Graf 3 **Kruhový graf - rozdelenie obyvateľov medzi krajinami V4 v roku 2016**



Zdroj: Vlastné spracovanie podľa údajov z www.kaggle.com

4.6 Histogram

Na obrázku nižšie (Obrázok 6) je zobrazený zdrojový kód na vykreslenie histogramu zobrazujúceho rozdelenie skúmaných krajín podľa priemernej dĺžky života. Na začiatku sa nachádza nastavenie minimálnej a maximálnej hodnoty v histograme a počtu stĺpcov. Hodnoty sú rozdelené na 4 intervaly. Prvý interval je od 50 do 60 rokov, druhý interval od 60 do 70 rokov, tretí interval od 70 do 80 rokov a štvrtý interval od 80 do 90 rokov. Potom je tu samotné volanie funkcie *histplot* nachádzajúcej sa v knižnici *seaborn* s použitím skôr nastavených parametrov. Ďalej je tu ešte zobrazenie popisov jednotlivých osí použitím funkcií *xlabel* a *ylabel* z knižnice *matplotlib*.

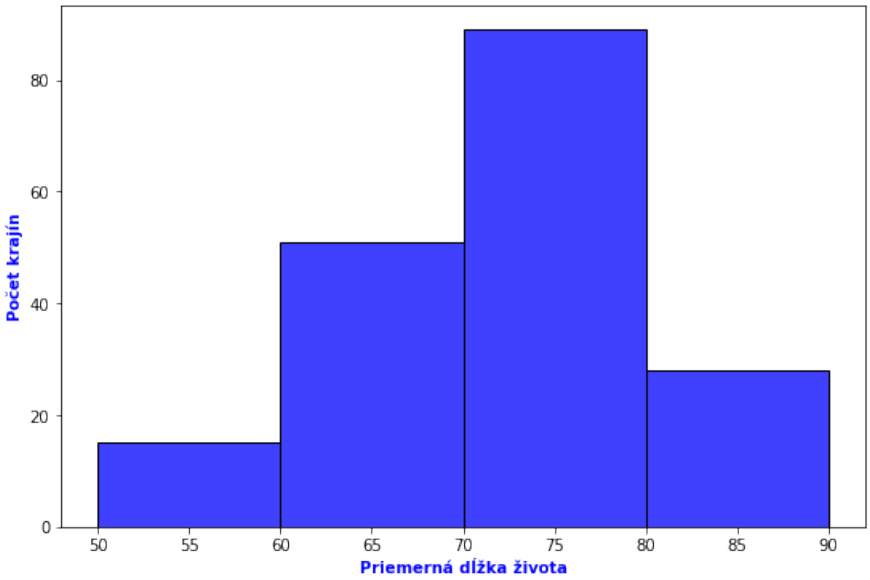
Obrázok 6 **Kód pre vytvorenie histogramu strednej dĺžkou života (2016)**

```
In [5]: plt.subplots(figsize = (9, 6))

        #Nastavenie minimálnej hodnoty, maximálnej hodnoty a počtu stĺpcov v histograme
        min_value = 50
        max_value = 90
        bins = 4

        #Vykreslenie histogramu
        sns.histplot(data = who_2016_df, x = "life_expect", bins = bins, binrange = (min_value, max_value), color = "blue")

        #Označenie osí x a y
        plt.xlabel("Priemerná dĺžka života", weight = "bold", color = "blue")
        plt.ylabel("Počet krajín", weight = "bold", color = "blue")

        plt.show()
```

Zdroj: Vlastné spracovanie

Ako je vidieť na grafe (Graf 4), výrazne najväčší počet krajín má priemernú dĺžku života v intervale 70 až 80 rokov. Zo všetkých skúmaných krajín to predstavuje takmer polovicu. Menej krajín je v intervale 60 až 70 rokov, ešte menej v intervale 80 až 90 rokov a zvyšné krajiny majú priemernú dĺžku života v intervale 50 až 60 rokov.

Graf 4 **Histogram - Priemerná dĺžka života v roku 2016**



Zdroj: Vlastné spracovanie podľa údajov z www.kaggle.com

4.7 Škatuľkový graf

Na obrázku nižšie (Obrázok 7) je zobrazený zdrojový kód, ktorý vykresľuje škatuľkové grafy vytvorené pre všetky regióny nachádzajúce sa v skúmaných údajoch, pričom každý z nich je vytvorený z údajov o priemernej dĺžke života v jednotlivých krajinách tohto regiónu v roku 2016. Na začiatku je vytvorený slovník s názvom *region_sk_dict*, v ktorom kľúče sú pôvodné názvy regiónov v anglickom jazyku, tak ako sa nachádzajú v zdrojových dátach a hodnoty sú slovenské preklady týchto názvov. V ďalšej časti kódu sú vytiahnuté údaje o hodnotách strednej dĺžky života v roku 2016 spolu s regiónom, v ktorom sa krajina s touto hodnotou nachádza a k jednotlivým regiónom je s použitím skôr vytvoreného slovníka pripojený aj slovenský názov daného regiónu. Ďalej sa tu nachádza nastavenie zobrazenia hodnôt aritmetických priemerov v jednotlivých škatuľkových grafoch, pričom sú tu nastavené ako biely kríž. Ďalej je tu volaná funkcia boxplot z knižnice *seaborn*, ktorá vykresľuje škatuľkové grafy a v jej parametroch sú nastavené hodnoty parametrov k jednotlivým škatuľkovým grafom ako slovenské názvy regiónov. Je tu použité nastavenie zobrazenia aritmetických priemerov v týchto grafoch s použitím vyššie vyvoreného slovníka *meanprops* a pomocou nastavenia parametra whis na (0, 100) je zabezpečené, aby každý z grafov zobrazoval hodnoty od minima po maximum v rámci daného regiónu.

Obrázok 7 **Kód - škatuľkové grafy so strednou dĺžkou života (2016)**

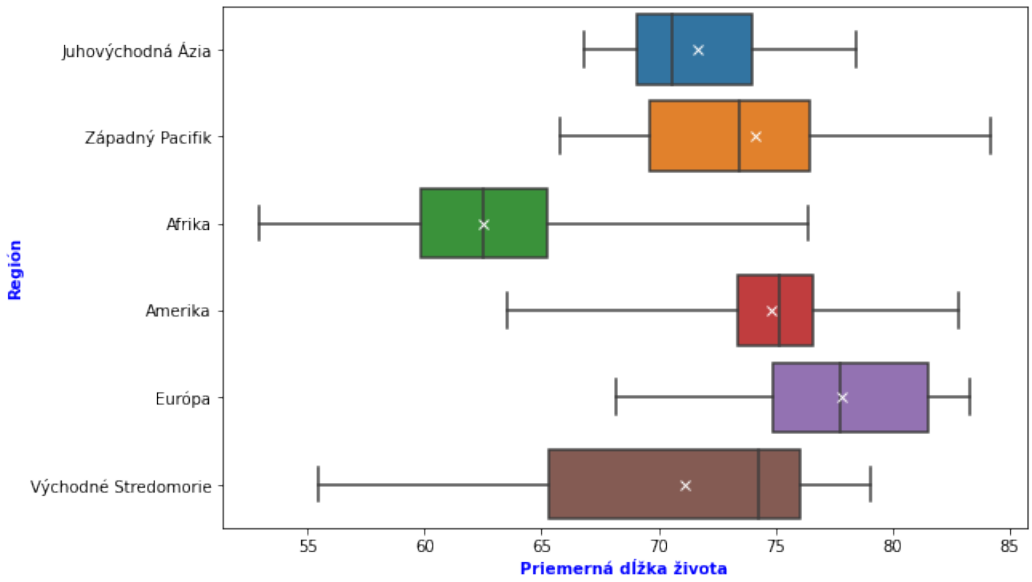
```
In [6]: plt.figure(figsize = (9, 6))

        #Zoznam slovenských názvov regiónov
        region_sk_dict = {"South-East Asia" : "Juhovýchodná Ázia",
                          "Western Pacific" : "Západný Pacifik",
                          "Africa" : "Afrika",
                          "Americas" : "Amerika",
                          "Europe" : "Európa",
                          "Eastern Mediterranean" : "Východné Stredomorie"}

        #Výber hodnôt strednej dĺžky života a pridanie informácie o slovenskom názve regiónu
        life_expect_df = who_2016_df[["region", "life_expect"]].copy()
        life_expect_df["region_sk"] = life_expect_df["region"].map(region_sk_dict)

        #Parametre zobrazenia priemeru v škatuľkových grafoch
        meanprops = {"marker" : "x", "markeredgecolor" : "white"}

        #Vykreslenie škatuľkových grafov
        sns.boxplot(data = life_expect_df, x = "life_expect", y = "region_sk",
                    order = region_sk_dict.values(), dodge = False, showmeans = True,
                    showfliers = False, meanprops = meanprops, whis = (0, 100))

        #Označenie osí x a y
        plt.xlabel("Priemerná dĺžka života", weight = "bold", color = "blue")
        plt.ylabel("Región", weight = "bold", color = "blue")

        plt.show()
```

Zdroj: Vlastné spracovanie

Z grafu (Graf 5) je viditeľné, že ľavý okraj škatuľkového grafu je najviac posunutý vľavo pre región Afrika, z čoho vyplýva, že najnižšia hodnota pre priemernú dĺžku života v krajine sa nachádza v regióne Afrika. Pravý okraj grafu sa nachádza najviac vpravo pre región Západný Pacifik, takže konkrétna krajina s najvyššou hodnotou priemernej dĺžky života sa nachádza v tomto regióne. Graf pre región Afrika je oproti ostatným regiónom posunutý výrazne doľava, pričom má jednoznačne najviac vľavo všetky hodnotené ukazovatele, takže hodnoty v tomto regióne sú jednoznačne najnižšie. Najvyššiu hodnotu minima, prvého kvartilu, mediánu, tretieho kvartilu a aritmetického priemeru má región Európa. Najväčšiu vzdialenosť medzi ľavým a pravým okrajom grafu je v grafe pre región Východné Stredomorie, takže údaje pre tento región majú najväčšie variačné rozpätie. Najmenšie variačné rozpätie majú hodnoty pre región Juhovýchodná Ázia. Najširší vnútorný obdĺžnik má škatuľkový graf zobrazujúci hodnoty pre región Východné Stredomorie, takže najväčšie kvartilové rozpätie majú hodnoty pre krajiny z tohto regiónu. Naopak, najužší vnútorný obdĺžnik je viditeľný v škatuľkovom grafe pre región Amerika, hodnoty pre priemernú dĺžku života majú teda najmenšie kvartilové rozpätie spomedzi regiónov.

Grafy pre regióny Juhovýchodná Ázia a Európa majú pravú časť vnútorného obdĺžnika výrazne väčšiu ako jeho ľavú časť, čo znamená, že rozdelenia s hodnotami pre tieto oblasti sú výrazne pravostranne zošikmené. Grafy pre Východné Stredomorie a Západný Pacifik majú výrazne väčšiu ľavú časť, takže rozdelenia pre tieto regióny sú výrazne ľavostranne zošikmené, pričom pri regióne Východné Stredomorie je toto zošikmenie výraznejšie.

Graf 5 **Škatuľkový graf - priemerná dĺžka života podľa oblasti (2016)**



Zdroj: Vlastné spracovanie podľa údajov z www.kaggle.com

Tabuľka zobrazená nižšie (Tabuľka 9) zobrazuje hodnoty minima, prvého kvartilu, mediánu, tretieho kvartilu, maxima a aritmetického priemeru pre rozdelenia vyjadrujúce strednú dĺžku života v jednotlivých oblastiach, ktoré sú graficky zobrazené v grafe vyššie (Graf 5) spoločne s hodnotami variačného rozptylu, kvartilového rozptylu a kvartilovej miery šikmosti, ktoré je z týchto hodnôt možné vypočítať. Všetky hodnoty v tabuľke sú zaokrúhlene na 2 desatinné miesta.

Tabuľka 9 **Hodnoty zobrazené v škatuľkových grafoch**

Región	min	Q_1^4	$\tilde{x}$	Q_3^4	max	$\bar{x}$	R	R_Q^4	S_Q
Juhovýchodná Ázia	66.80	69.07	70.57	74.00	78.42	71.64	11.61	4.93	0.39
Západný Pacifik	65.79	69.60	73.43	76.43	84.17	74.13	18.38	6.83	-0.12
Afrika	52.94	59.81	62.50	65.24	76.36	62.53	23.43	5.43	0.01
Amerika	63.51	73.40	75.17	76.56	82.81	74.80	19.30	3.16	-0.12
Európa	68.16	74.88	77.76	81.52	83.26	77.82	15.10	6.64	0.13
Východné Stredomorie	55.45	65.31	74.28	76.05	79.06	71.15	23.62	10.73	-0.67

Zdroj: Vlastné spracovanie podľa údajov z www.kaggle.com

4.8 Kartogram

4.8.1 Geografické údaje

Pre vykreslenie jednotlivých štátov na mape slúžia údaje získané z adresy https://datahub.io/core/geo-countries. Je použitý súbor countries.geojson, ktorý je vo formáte GeoJSON a obsahuje polygóny jednotlivých krajín. Stĺpce, ktoré sú z tohto súboru ďalej použité, sú zobrazené v tabuľke nižšie (Tabuľka 10). [24]

Tabuľka 10 **Zoznam použitých premenných súboru countries.geojson**

Názov premennej	**Popis**
ISO_A3	kód krajiny vo formáte ISO3166-1-Alpha-3
Geometry	polygón krajiny

Zdroj: Vlastné spracovanie podľa údajov z datahub.io

Nižšie je zobrazený zdrojový kód pre zobrazenie kartogramu (Obrázok 8), ktorý zobrazuje farbu krajiny na svetovej mape podľa priemernej dĺžky života v danej krajine v roku 2016. Na začiatku je vytvorený slovník pre zobrazenie krajín a území s chýbajúcimi údajmi bledosivou farbou. Ďalšia časť zdrojového kódu vytvára slovník pre použitie nastavenia legendy grafu, konkrétne je tu nastavený popis legendy a jej horizontálna orientácia. Táto mapa je ďalej vytvorená pomocou funkcie *plot*, ktorá sa nachádza v knižnici *geopandas*. Ako jej parametre sú použité slovníky *missing_kwds* a *legend_kwds* vytvorené vyššie a ako stĺpec, podľa ktorého sa zobrazí farba na mape je v tomto zdrojovom kóde vybraný stĺpec *life_expect*, ktorý obsahuje informácie o priemernej dĺžke života v jednotlivých krajinách. V tomto prípade sa farebná mapa v kóde nenastavuje, preto je pri vykreslení mapy použitá predvolená farebná mapa pre mapy v knižnici *geopandas*, ktorá z modrej farby postupne prechádza do žltej. Pre krajiny alebo oblasti s chýbajúcimi údajmi je použitá bledosivá farba, ako bolo nastavené v zdrojovom kóde. Na osi *x* na mape sú znázornené hodnoty pre geografickú dĺžku a na osi *y* sú zobrazené hodnoty pre geografickú šírku.

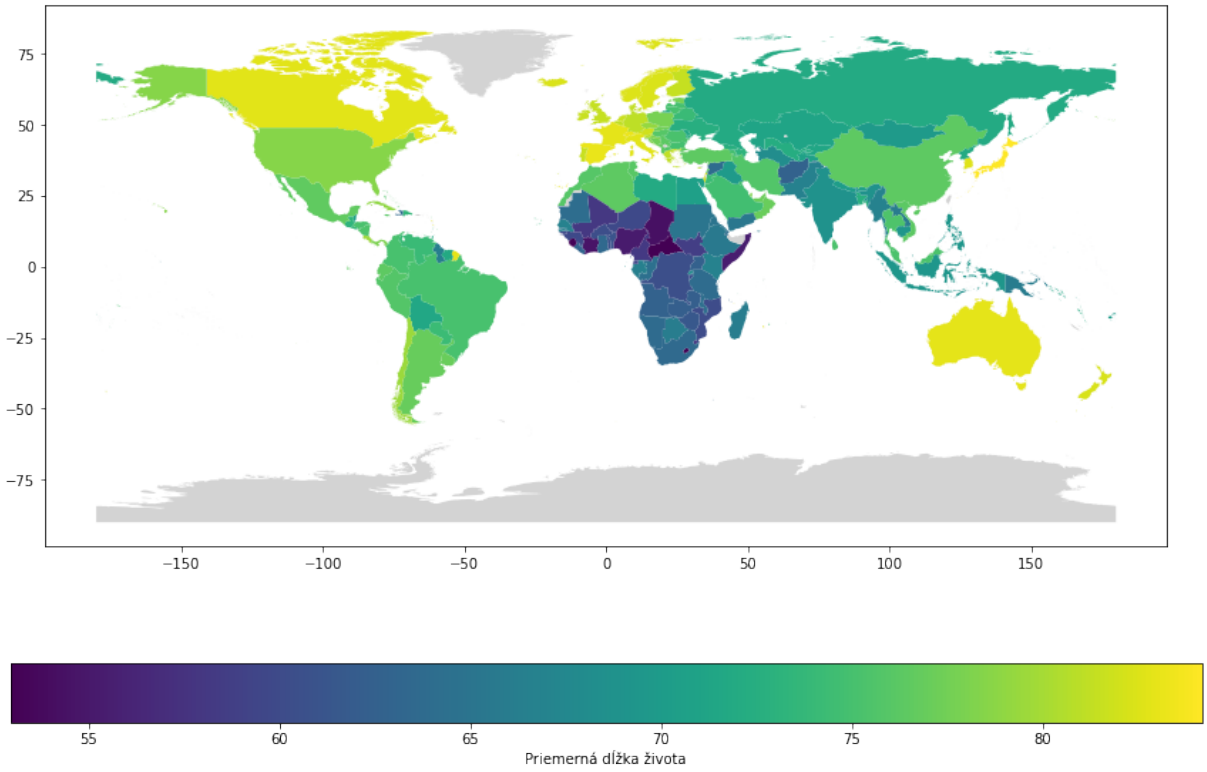
Obrázok 8 **Kód pre vykreslenie kartogramu s priemernou dĺžku života (2016)**

```
In [7]: #Parametre pre zobrazenie chýbajúcich údajov
        missing_kwds = {"color" : "lightgrey"}

        #Parametre pre zobrazenie legendy
        legend_kwds = {"label" : "Priemerná dĺžka života", "orientation" : "horizontal"}

        #Vykreslenie kartogramu
        map_data_df.plot(column = "life_expect", legend = True, figsize = (15, 10),
                         missing_kwds = missing_kwds, legend_kwds = legend_kwds)

        plt.show()
```

Zdroj: Vlastné spracovanie

Ako je vidieť na mape (Graf 6), najvyššie hodnoty priemernej dĺžky života zobrazené odtieňmi približujúcimi sa k žltej farbe sú na mape viditeľné v Kanade, v západnej Európe, v časti východnej Ázie, v Austrálii a na Novom Zélande. Viac zelené oblasti s menšou priemernou dĺžkou života sú potom s rôznymi odtieňmi zelenej Amerika okrem Kanady, severná Afrika a väčšina Ázie. Jednoznačne najvýraznejšie zastúpenie krajín s nízkou priemernou dĺžkou života s odtieňmi farby približujúcimi sa k modrej farbe je viditeľné v Afrike s výnimkou jej severnej časti.

Graf 6 **Kartogram - priemerná dĺžka života v roku 2016**



Zdroj: Vlastné spracovanie podľa údajov z www.kaggle.com a datahub.io

Na obrázku nižšie (Obrázok 9) je zobrazený zdrojový kód pre vytvorenie kartogramu, ktorý zobrazuje farbu krajiny na svetovej mape podľa údaju z roku 2016 o percentuálnom podiele obyvateľov danej krajiny, ktorí mali aspoň základný prístup k pitnej vode.

Na začiatku je vytvorený slovník pre zobrazenie krajín a území s chýbajúcimi údajmi tmavosivou farbou. Ďalšia časť vytvára slovník pre použitie nastavenia legendy grafu, konkrétne je tu nastavený popis legendy a jej horizontálna orientácia. Ďalej je tu nastavenie farebnej mapy pre použitie pri zobrazení mapy. Ako farebná mapa pre tento kartogram bola použitá farebná mapa *plasma_r*, čo reprezentuje farebnú mapu *plasma* nachádzajúcu sa v rozhraní *matplotlib.pyplot* v obrátenom poradí. Pôvodná farebná mapa *plasma* obsahuje prechod z tmavo fialovej farby blížiacej sa k modrej, cez ružovú postupne do žltej farby, pri obrátenej farebnej mape *plasma_r* je to opačne. Táto mapa je ďalej vytvorená pomocou funkcie *plot*, ktorá sa nachádza v knižnici *geopandas*. Ako jej parametre sú použité slovníky vytvorené vyššie, skôr nastavená farebná mapa a ako stĺpec, podľa ktorého sa zobrazí farba na mape je tu vybratý stĺpec *basic_water*, ktorý obsahuje informácie o percentuálnom podiele obyvateľov s prístupom k pitnej vode v jednotlivých krajinách. Na osi *x* na mape sú znázornené hodnoty pre geografickú dĺžku a na osi *y* sú zobrazené hodnoty pre geografickú šírku.

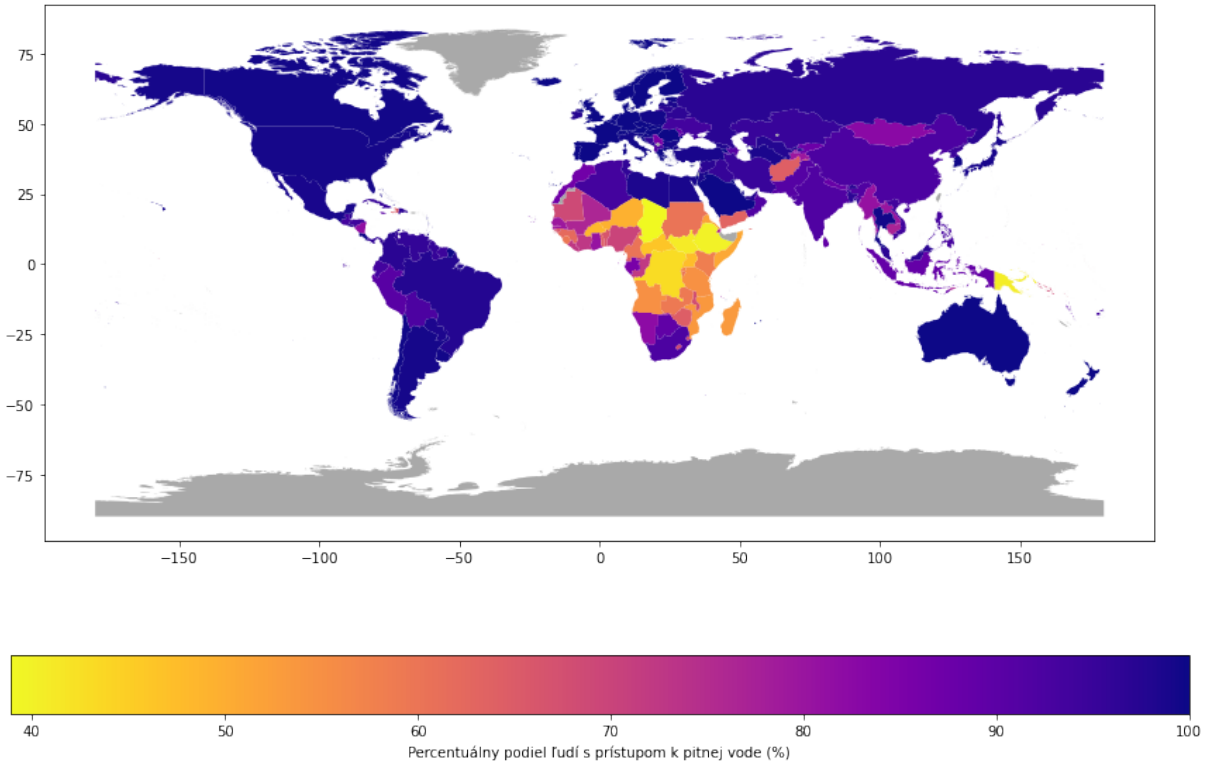
Obrázok 9 **Kód - mapa pre podiel obyvateľov s prístupom k pitnej vode (2016)**

```
In [8]: basic_water_str = "Percentuálny podiel ľudí s prístupom k pitnej vode (%)"

        #Parametre pre zobrazenie chýbajúcich údajov
        missing_kwds = {"color" : "darkgrey"}

        #Parametre pre zobrazenie legendy
        legend_kwds = {"label" : basic_water_str, "orientation" : "horizontal"}

        #Výber farebnej mapy
        colormap = "plasma_r"

        #Vykreslenie kartogramu
        map_data_df.plot(column = "basic_water", legend = True, figsize = (15, 10),
                         missing_kwds = missing_kwds, legend_kwds = legend_kwds, cmap = colormap)

        plt.show()
```

Zdroj: Vlastné spracovanie

Ako je vidieť na mape (Graf 7), medzi krajiny s vysokým podielom obyvateľov s prístupom k pitnej vode v roku 2016 patrila väčšina štátov Ameriky s výnimkou niektorých krajín v Južnej Amerike s o niečo nižším podielom ľudí s prístupom k pitnej vode. Ďalej tam patrila väčšina Európy s výnimkou niektorých krajín na jej východe, kde bol podiel obyvateľov s prístupom k pitnej vode o niečo nižší, väčšina severnej a časť južnej Afriky a takisto Austrália a Nový Zéland. Výrazne nižšie percentuálne podiely sa vyskytovali v niektorých ázijských krajinách. Viditeľne najvýraznejší podiel krajín s nízkym podielov obyvateľov s prístupom k pitnej vode sa vyskytoval vo väčšine Afriky s výnimkou jej severnej a južnej časti, pričom na mape sa vo veľa prípadoch blíži odtieň farby k žltej, čo predstavuje najnižší podiel takýchto obyvateľov v rámci mapy blížiaci sa k 40 percentám.

Graf 7 **Kartogram - percentuálny podiel ľudí s prístupom k pitnej vode v roku 2016**



Zdroj: Vlastné spracovanie podľa údajov z www.kaggle.com a datahub.io

4.9 Tepelná mapa

Na obrázku nižšie (Obrázok 10) je zobrazený zdrojový kód pre vytvorenie tepelnej mapy, ktorá zobrazuje priemernú mesačnú teplotu v Bratislave v jednotlivých mesiacoch v rokoch 2012 až 2019. Na začiatku je definovaná funkcia *celsius_to_fahrenheit*, ktorá na vstupe dostáva hodnotu v stupňoch Fahrenheita a prevedie ju na stupne Celsia. Ďalej sú tu zo vstupného datasetu vytiahnuté záznamy pre mesto Bratislava do nového objektu typu *DataFrame* s názvom *weather_ba_df*, kde sú ďalej hodnoty usporiadané podľa roka, mesiaca a dňa. Do tohto *DataFramu* je ďalej pridaný stĺpec *celsius*, ktorý obsahuje teploty v stupňoch Celsia a tieto hodnoty sú prepočítané z pôvodných hodnôt v stupňoch Fahrenheita v stĺpci *AvgTemperature* pomocou skôr vytvorenej funkcie a sú tu dopočítané chýbajúce hodnoty pomocou funkcie *interpolate* z knižnice *pandas*.

Z týchto hodnôt sú ďalej vytiahnuté hodnoty pre roky 2012 až 2019 a nad tým je zavolaná funkcia *pivot_table* z objektu *DataFrame*, ktorá vytvára výslednú tabuľku, kde sú mesiace ako riadky a roky ako stĺpce, pričom hodnoty v *DataFrame* predstavujú aritmetický priemer hodnôt z pôvodného datasetu pre príslušný rok a mesiac. Nad týmto *DataFramom* je ďalej zavolaná funkcia *heatmap* z knižnice *seaborn*, ktorá vykresľuje prvky tohto *DataFramu* do tepelnej mapy. Ako farebná mapa je tu použitá farebná mapa *coolwarm*, ktorá sa nachádza v knižnici *matplotlib*. Hodnoty sa tu zaokrúhľujú na 1 desatinné miesto. Na konci sú zavolané funkcie *xlabel* a *ylabel* z knižnice *matplotlib*, ktoré nastavujú popisy jednotlivých osí na Rok a Mesiac.

Obrázok 10 **Zdrojový kód pre vykreslenie tepelnej mapy**

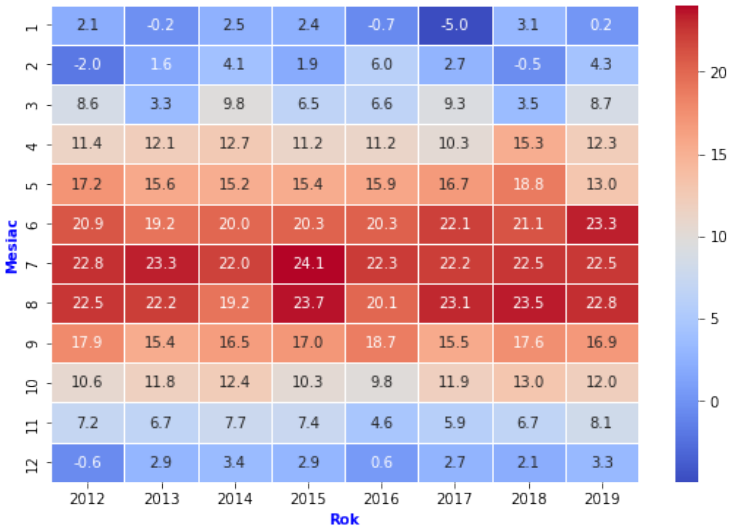
```
In [11]: #Funkcia na prepočet teploty zo stupňov Fahrenheita na stupne Celsia
def celsius_to_fahrenheit(fahrenheit):
    celsius = (fahrenheit - 32) * 5 / 9
    return celsius

#Vytiahnutie hodnôt o teplote v Bratislave a nahradenie chýbajúcich hodnôt
weather_ba_df = weather_df[(weather_df["City"] == "Bratislava")].copy()
weather_ba_df.sort_values(by = ["Year", "Month", "Day"], inplace = True)
weather_ba_df.replace(-99, np.nan, inplace = True)
weather_ba_df.interpolate(inplace = True)

#Prevod teploty do stupňov Celsia a zoradenie hodnôt
weather_ba_df["celsius"] = weather_ba_df["AvgTemperature"].apply(celsius_to_fahrenheit)


#Výpočet pivotnej tabuľky pre zobrazenie
temp_avg_df = weather_ba_df[weather_ba_df["Year"].isin(range(2012, 2020))].pivot_table(values = "celsius", index = "Month",
                                                                                       columns = "Year", aggfunc = "mean")

#Vytvorenie možnosti zobrazenia viacerých grafov súčasne
fig, ax = plt.subplots(figsize = (9, 6))

#Výber farebnej mapy
colormap = "coolwarm"

#Vykreslenie tepelnej mapy
sns.heatmap(temp_avg_df, annot = True, fmt = "1.1f",
            linewidths = ".5", ax = ax, cmap = colormap)

#Zobrazenie legendy
plt.xlabel("Rok", weight = "bold", color = "blue")
plt.ylabel("Mesiac", weight = "bold", color = "blue")

plt.show()
```

Zdroj: Vlastné spracovanie

Na tepelnej mape (Graf 8) sú viditeľné jednoznačne vymedzené ročné obdobia v danom regióne, kde sa rozdiely priemerných mesačných teplôt medzi letom a zimou pohybujú okolo 20 stupňov Celsia. Teploty v januári sa pohybovali väčšinu sledovaného obdobia v rozmedzí hodnôt -1 až 3 stupne, výnimkou bol rok 2017, kedy január bol jednoznačne najchladnejší mesiac v sledovanom období. Februárové teploty zaznamenali výraznejšie výkyvy v rokoch 2012 a 2018, kedy bola priemerná februárová teplota pod nulou na rozdiel od ostatných sledovaných rokov a v roku 2016, kedy bol naopak február výrazne teplejší ako v ostatných rokoch. Ohľadne jarných mesiacov boli apríl a máj najteplejšie v roku 2018, ktoré nasledovali po chladnom marci, vysoké májové teploty boli zaznamenané aj v rokoch 2012, 2014, 2017 a 2019. Mesiace júl a august sa vyznačovali veľmi vysokými teplotami v roku 2015, rok predtým, v roku 2014, boli naopak tieto mesiace veľmi chladné. Vysokými teplotami počas celého jesenného obdobia sa vyznačovali oba dané roky 2014 a 2015. December bol najchladnejší v roku 2012, keď nasledoval po teplom lete a jeseni a bol zároveň jediným decembrom v sledovanom období s mínusovou priemernou teplotou.

Celkovo najteplejší mesiac počas daného sledovaného obdobia bol júl 2015 s priemernou teplotou 24.1 stupňa Celsia a najchladnejší bol január 2017 s priemernou teplotou -5.0 stupňa Celsia.

Graf 8 **Tepelná mapa - priemerná mesačná teplota v Bratislave**



Zdroj: Vlastné spracovanie podľa údajov z www.kaggle.com

5 Diskusia

V podkapitole 4.3 je zobrazený spojnicový graf, ktorý zobrazuje vývoj hodnôt priemernej spotreby alkoholu na osobu v krajinách Vyšehradskej štvorky v rokoch 2000 až 2016. Ukazuje to jedno z možných použití funkcie plot nachádzajúcej sa v knižnici *matplotlib*, ktorá je súčasťou programovacieho jazyka Python, kde je viditeľný vývoj daného ukazovateľa v čase pre jednotlivé krajiny. Nižšie pod daným grafom sú uvedené základné charakteristiky časových radov vykreslených v danom grafe pre jednotlivé krajiny, ktoré umožňujú bližšie charakterizovať a opísať tieto rady a ďalej sú z nich vypočítané priemery, ktoré charakterizujú dané časové rady výstižnejšie a tieto sú potom ešte zhrnuté v ďalšej tabuľke.

V podkapitole 4.4 je zobrazený stĺpcový graf, ktorý zobrazuje 10 krajín s najväčšou spotrebou čistého alkoholu na osobu v roku 2016 spomedzi hodnotených krajín. Hodnoty spotreby v týchto krajinách sú zobrazené na pravom okraji daných stĺpcov bielou farbou boldom. Takéto zobrazenie daných hodnôt zabezpečí dobrú viditeľnosť týchto ukazovateľov v rámci tohto grafu a zobrazenie hodnôt priamo na jednotlivých farebných stĺpcoch zabezpečí jednoznačné a rýchle rozoznanie príslušnosti hodnôt k jednotlivým krajinám.

V podkapitole 4.5 je vykreslený kruhový graf zobrazujúci percentuálne rozdelenie obyvateľov Vyšehradskej štvorky medzi jednotlivými krajinami v roku 2016. Takéto zobrazenie ponúka vizuálnu reprezentáciu pomeru počtu obyvateľov v jednotlivých krajinách k celkovému súčtu počtov obyvateľov daných krajín.

V podkapitole 4.6 sa nachádza histogram, ktorý rozdeľuje skúmané krajiny podľa priemernej dĺžky života v roku 2016 do jednotlivých intervalov.

Podkapitola 4.7 obsahuje zobrazenie škatuľkových grafov priemernej dĺžky života v roku 2016 pre jednotlivé rozdelenia vytvorené podľa jednotlivých regiónov definovaných v rámci skúmaných dát. Škatuľkové grafy sú farebne odlíšené pre lepšiu rozpoznateľnosť a sú označené názvom daného regiónu, ktorý sa nachádza na osi y z dôvodu jednoznačného priradenia každého zo škatuľkových grafov k danému regiónu. Aritmetický priemer týchto rozdelení je vo všetkých škatuľkových grafoch zobrazený bielou farbou, čo zabezpečuje dobré vizuálne odlíšenie od farebných oblastí týchto grafov ako aj od čiernych čiar ohraničujúcich

jednotlivé oblasti grafov. Pod týmto grafom sa nachádza tabuľka, ktorá pre jednotlivé regióny uvádza hodnoty vykreslené v škatuľkovom grafe vyššie, ktorými sú minimum, prvý kvartil, medián, tretí kvartil, maximum a aritmetický priemer hodnôt pre tento región a takisto sú v nej uvedené aj hodnoty variačného rozpätia, kvartilového rozpätia a kvartilovej miery šikmosti, ktoré umožňujú lepšie charakterizovať dané rozdelenia.

V podkapitole 4.8 sú zobrazené 2 kartogramy, ktoré zobrazujú rozdelenie skúmaných krajín podľa priemernej dĺžky života v týchto krajinách v roku 2016 a podľa percentuálneho podielu obyvateľov, ktorí mali prístup k pitnej vode v roku 2016. Jednotlivé štáty sú na oboch týchto mapách farebne odlíšené podľa hodnoty týchto premenných, pričom pod týmito mapami je vždy zobrazená farebná mapa, ktorá umožňuje priradenie farby jednotlivých štátov k hodnotám daných ukazovateľov. Krajiny alebo oblasti, ktoré sa nachádzajú na mape ale nie sú medzi skúmanými krajinami, sú vyznačené sivou farbou, ktorej odtieň je v oboch mapách zvolený podľa použitých farieb pre krajiny so známou hodnotou daného ukazovateľa tak, aby bolo odlíšenie týchto krajín dobre viditeľné a rozpoznateľné.

V podkapitole 4.9 sa nachádza tepelná mapa, ktorá zobrazuje priemerné mesačné teploty v Bratislave vo všetkých mesiacoch rokov 2012 až 2019. Riadky tejto tepelnej mapy tvoria jednotlivé mesiace a stĺpce sú tvorené rokmi. Takéto zobrazenie umožňuje vidieť v danom zobrazení porovnanie hodnôt podľa jednotlivých mesiacov a takisto aj podľa roka v rámci skúmaného obdobia.

Záver

Táto práca mala za cieľ ukázať možnosti vizualizácie rôznych typov dát s použitím základných, ale aj pokročilejších grafických zobrazení s použitím viacerých dostupných a často používaných knižníc programovacieho jazyka Python, pričom na spracovanie a vizualizáciu dát boli v tejto práci použité knižnice *numpy*, *pandas*, *matplotlib*, *seaborn* a *geopandas*.

Práca zobrazuje viaceré základné aj pokročilejšie grafické zobrazenia vytvorené s použitím dát rôzneho typu z webovej stránky www.kaggle.com, pomocných dát obsahujúcich slovenské názvy štátov nachádzajúcich sa na webovej stránke sk.wikipedia.org a geografických dát z webovej stránky datahub.io pomocou programovacieho jazyka Python a vývojového prostredia Jupyter spolu s ukážkami zdrojových kódov, pomocou ktorých boli tieto grafy zobrazené a ktoré sú najvýznamnejšou súčasťou výsledkov práce.

Presnosť údajov vyobrazených v týchto zobrazeniach je daná predovšetkým presnosťou informácií nachádzajúcich sa vo vstupných údajoch, ktoré boli v tejto práci použité, čiastočne môžu byť ovplyvnené aj zaokrúhlením alebo nedostupnosťou niektorých chýbajúcich údajov, najmä pri výpočte priemerných mesačných teplôt v Bratislave, kde boli niektoré chýbajúce denné hodnoty dopočítané, počet takto dopočítaných údajov však bol pomerne nízky. Zdroje použitých údajov sú takisto uvedené v časti o výsledkoch práce, takže je možné si tieto vstupné dáta overiť.

Tieto zobrazenia vždy znázorňujú legendu, ktorej forma je zvolená podľa jednotlivého zobrazenia pre zabezpečenie čo najlepšej zrozumiteľnosti. V zobrazeniach, kde sú znázornené 2 rôzne premenné na osiach x a y, sú názvy týchto premenných zobrazené modrou farbou boldom pre jednoznačné odlíšenie od okolitého textu, ktorý má čiernu farbu. V ostatných zobrazeniach boli zvolené iné spôsoby zobrazenia, čo predstavovalo napríklad zobrazenie legendy v rohu daného grafu alebo zobrazenie pod farebnou mapou pri zobrazeniach, kde je znázornená mapa sveta.

V rámci výsledkov práce sú viditeľné viaceré vybrané druhy grafických zobrazení, ktoré je možné vytvoriť pomocou knižníc programovacieho jazyka Python, ktoré boli v tejto práci použité. Tu je ukázaná práca s knižnicami a niektoré vybrané funkčnosti z veľkého množstva možností, ktoré tieto knižnice ponúkajú. Výsledky práce obsahujú aj zdrojové kódy, pomocou ktorých boli tieto zobrazenia vytvorené, čo umožní do budúcnosti vytvárať rovnaké alebo podobné zobrazenia s použitím dát, ktoré takéto zobrazenia umožnia.

Zoznam literatúry

[1] MCKINNEY, Wes. *Python for Data Analysis.* Druhé vydanie. Sevastopoľ : O'Reilly Media, Inc., 2017. ISBN 978-1-491-95766-0

[2] The official home of the Python Programming Language. [Online] [Dátum: 28. Apríl 2022.] https://www.python.org.

[3] Project Jupyter. [Online] [Dátum: 1. Apríl 2022.] https://www.jupyter.org.

[4] Anaconda. [Online] [Dátum: 1. Apríl 2022.] https://www.anaconda.com.

[5] KOTLEBOVÁ, Eva a kol. *Štatistika pre bakalárov v praxi.* Prvé vydanie. Bratislava : Vydavateľstvo EKONÓM, 2017. 318 strán. ISBN 978-80-225-4366-8

[6] Sloupcový graf – Wikipedie. [Online] 30. 03 2022. [Dátum: 1. Apríl 2022.] https://cs.wikipedia.org/wiki/Sloupcový_graf.

[7] Heat map - Wikipedia. [Online] 10. 01 2022. [Dátum: 1. Apríl 2020.] https://en.wikipedia.org/wiki/Heat_map.

[8] GEMIGNANI, Zach a kol. *Efektivní analýza a využití dat.* Prvé vydanie. Brno : Computer Press, 2015. ISBN 978-80-251-4571-5.

[9] SORANSON, Oliver. *Python Analysis.* Prvé vydanie. 2019. 170 strán.

ISBN 9781706623588.

[10] PAJANKAR, Ashwin. *Practical Python Data Visualization.* Druhé vydanie. Berlín : Springer, 2020. 160 strán. ISBN 978-1-4842-6454-6.

[11] Pecinovský, Rudolf. *Python - Kompletní příručka jazyka pro verzi 3.9.* Prvé vydanie. Praha : Grada Publishing, a.s., 2020. 480 strán. ISBN 978-80-271-1851-9.

[12] SARGENT, Thomas J. – STACHURSKI, John. Quantitative Economics with

Python [Dátum: 2. Február 2021.]

[13] Matplotlib - Visualization with Python**.** [Online] [Dátum: 15. Apríl 2022.]. https://matplotlib.org.

[14] Pandas - Python Data Analysis Library**.** [Online] [Dátum: 16. Apríl 2022.]. https://pandas.pydata.org.

[15] NAVLANI, Avinash – FANDANGO, Armando – IDRIS, Ivan. *Python Data Analysis.* Tretie vydanie. Birmingham : Packt Publishing Ltd, 2021. 478 strán.

ISBN 978-1-78995-524-8

[16] TOMS, Silas – CRICKARD, Paul - VAN REES, Eric. *Mastering Geospatial Analysis with Python.* Prvé vydanie. Birmingham : Packt Publishing Ltd, 2018. 440 strán.

ISBN 978-1-78829-333-4.

[17] Seaborn: statistical data visualization**.** [Online] [Dátum: 17. Apríl 2022.]. https://seaborn.pydata.org.

[18] GeoPandas. [Online] [Dátum: 17. Apríl 2022.]. https://geopandas.org.

[19] Kaggle: Your Home for Data Science [Online] [Dátum: 29. Apríl 2022.]. https://www.kaggle.com.

[20] Matplotlib Tutorial | Python Matplotlib Library with Examples | Edureka [Online] [Dátum: 17. Apríl 2022.]. https://www.edureka.co/blog/python-matplotlib-tutorial.

[21] Udemy [Online] [Dátum: 21. Apríl 2022.]. https://www.udemy.com.

[22] Python (programovací jazyk) – Wikipédia. [Online] 08. 03. 2022. [Dátum: 27. Apríl 2022.] https://sk.wikipedia.org/wiki/Python_(programovací_jazyk).

[23] NumPy. [Online] [Dátum: 29. Apríl 2022.]. https://numpy.org.

[24] DataHub [Online] [Dátum: 29. Apríl 2022.]. https://datahub.io/core/geo-countries.

[25] ISO 3166-1 – Wikipédia. [Online] 22. 02. 2022. [Dátum: 29. Apríl 2022.] https://sk.wikipedia.org/wiki/ISO_3166-1.

[26] Python – Wikipedie. [Online] 08. 03 2022. [Dátum: 29. Apríl 2022.] https://cs.wikipedia.org/wiki/Python.