

Univerzita Palackého v Olomouci
Přírodovědecká fakulta
Katedra geoinformatiky

**TVORBA ARCGIS NOTEBOOKS PRO
ZPRACOVÁNÍ DAT EVROPSKÝCH MĚST
A STÁTŮ**

Diplomová práce

Bc. Adam TÓTH

Vedoucí práce doc. Ing. Zdena DOBEŠOVÁ, Ph.D.

Olomouc 2024

Geoinformatika a kartografie

ANOTÁCIA

Cieľom práce je vytvoriť ArcGIS Notebooky, každý zameraný na inú tému, ktoré spracovávajú dáta z územia európskych štátov a miest. ArcGIS Notebook je implementáciou Jupyter Notebooku do ArcGIS prostredia. Cieľom notebookov je komplexne oboznámiť užívateľa s vybranou témou a spracovanými dátami formou “data storytelling”. Ten vzniká kombináciou vysvetľujúceho textu, programového kódu a vizualizácií vo forme tabuliek, grafov a máp, pričom táto kombinácia a striedanie buniek textu s bunkami kódu tvorí základ Jupyter Notebooku. Vybranými témami sú: kolokačná analýza využitia územia v mestskom prostredí, hustota dopravnej siete, automatizovaný prístup k dátam z Eurostatu, spracovanie satelitných snímok za účelom odhadu rozlohy ľadovca a tvorba máp pomocou Python knižníc. Notebooky boli vytvorené v programe Visual Studio Code a v ArcGIS Pro a sú napísané v anglickom jazyku. Programovacím jazykom je Python. Počet vytvorených notebookov, ktoré sú výstupom tejto diplomovej práce, je osem. Tri z nich sú úvodom do jazyka Python, knižnice ArcPy a práce s Jupyter Notebookom vo Visual Studio Code a v ArcGIS Pro. Tieto tri notebooky boli testované vo výuke na Katedre geoinformatiky PrF UP.

KLÚČOVÉ SLOVÁ

ArcGIS Notebook; Jupyter Notebook; Python; európske dáta

Počet strán práce: 81

Počet príloh: 2 (z toho 1 voľná a 1 elektronická)

ANOTATION

The aim of this thesis is to create ArcGIS Notebooks, with each dedicated to a different topic and processing data from the territory of European countries and cities. The ArcGIS Notebook is an implementation of the Jupyter Notebook in an ArcGIS environment. The aim of the notebooks is to comprehensively familiarise the user with the selected topic and the processed data in the form of "data storytelling". This is created by combining explanatory text, programming code and visualizations in the form of tables, graphs, and maps; this combination and the alternation of cells of text with cells of code forms the basis of the Jupyter Notebook. Selected topics include: colocation analysis of land use in urban environment, density of the transport network, automated access to Eurostat data, processing satellite imagery to estimate glacier extent and map creation using Python libraries. The notebooks were created in Visual Studio Code and ArcGIS Pro and are written in English. The programming language is Python. The number of notebooks created, which are the output of this thesis, is eight. Three of them are an introduction to the Python language, the ArcPy library, and working with the Jupyter Notebook in Visual Studio Code and ArcGIS Pro. These three notebooks have been tested in the teaching process at the Department of Geoinformatics at the Faculty of Science of Palacký University in Olomouc.

KEYWORDS

ArcGIS Notebook; Jupyter Notebook; Python; European data

Number of pages: 81

Number of appendixes: 2

Prehlasujem, že

- diplomovú prácu, vrátane príloh, som vypracoval samostatne a uviedol som všetky použité podklady a literatúru. Notebook Hustota dopravnej siete nadväzuje skript z bakalárskej práce Možnosti výpočtu charakteristik dopravní sítě států a měst (Tóth 2022). Notebook Rozloha řadovca vychádza zo semestrálnej práce GeoPython: výpočet rozlohy řadovca (Potočiar a Tóth 2023). Notebook Mapovacie knižnice je inšpirovaný článkom 6 python libraries to make beautiful maps (Rožanov 2024),

- som si vedomý, že na moju diplomovú prácu sa plne vzťahuje zákon č.121/2000 Zb. - autorský zákon, hlavne § 35 – využitie diela v rámci občianskych a náboženských obradov, v rámci školských predstavení a využitie diela školského a § 60 – školské dielo,

- beriem na vedomie, že Univerzita Palackého v Olomouci (ďalej UP Olomouc) má právo nezárobkovo, pre svoju vnútornú potrebu, diplomovú prácu používať (§ 35 ods. 3),

- súhlasím, že údaje o mojej diplomovej práci budú zverejnené v Študijnom informačnom systéme UP,

- v prípade záujmu UP Olomouc uzavriem licenčnú zmluvu s oprávnením použiť výsledky a výstupy mojej diplomovej práce v rozsahu § 12 ods. 4 autorského zákona,

- použiť výsledky a výstupy mojej diplomovej práce alebo poskytnúť licenciu k jej využitiu môžem len so súhlasom UP Olomouc, ktorá je oprávnená v takomto prípade od mňa požadovať primeraný príspevok na úhradu nákladov, ktoré boli UP Olomouc na vytvorenie diela vynaložené (až do ich skutočnej výšky).

Ďakujem vedúcej tejto diplomovej práce doc. Ing. Zdene Dobešovej, Ph.D. za odborné vedenie, konzultácie a pripomienky pri vypracovaní práce. Ďakujem mojej rodine, že ma podporovala v štúdiu. Ďakujem mojej snúbenici Janke za jej lásku, dôveru a oporu nielen v čase písania tejto diplomovej práce.

UNIVERZITA PALACKÉHO V OLOMOUCI

Přírodovědecká fakulta
Akademický rok: 2022/2023

ZADÁNÍ DIPLOMOVÉ PRÁCE

(projektu, uměleckého díla, uměleckého výkonu)

Jméno a příjmení: Bc. Adam TÓTH
Osobní číslo: R220654
Studijní program: N0532A330009 Geoinformatika a kartografie
Téma práce: Tvorba ArcGIS Notebooks pro zpracování dat evropských měst a států
Zadávající katedra: Katedra geoinformatiky

Zásady pro vypracování

Cílem práce je vytvořit ve formátu ArcGIS Notebook vhodná data-storytelling týkajících se evropských měst a států. Student vybere vhodná témata analýzy jako je doprava, využití území a další, která

formou ArcGIS Notebooku analyzuje, doplní komentáři, tabulkami, grafy a mapami. ArcGIS Notebooks budou sloužit k zpracování dat formou "dat s příběhem" tak, aby uživatel formou doprovodného textu a automatizovaného spouštění dílčích úseků Python kódu byl komplexně seznámen s daty a výsledky prostorových analýz pro vybraná témata evropských měst a států.

Text práce student zpracuje v souladu se závaznou šablonou pro kvalifikační práce KGI. O diplomové práci student vytvoří webovou stránku a poster. Celou práci (text, přílohy, výstupy, zdrojová a vytvořená data, poster a web) odevzdá student v digitální podobě na datové úložiště katedry. Do evidence STAG student odevzdá úplný text práce s přílohami, které určí vedoucí práce. Fyzicky student odevzdá výtisk posteru ve formátu A2 a přílohy určené vedoucím práce.

Rozsah pracovní zprávy: max. 50 stran
Rozsah grafických prací: dle potřeby
Forma zpracování diplomové práce: elektronická
Jazyk zpracování: Slovenština

Seznam doporučené literatury:

COPERNICUS PROGRAMME, 2023. Urban Atlas [online]. <https://land.copernicus.eu/local/urban-atlas>
Esri. Notebooks in ArcGIS Pro [online]. <https://pro.arcgis.com/en/pro-app/latest/arcpy/get-started/pro-notebooks.htm>
EUROSTAT. 2023. Eurostat database [online]. <https://ec.europa.eu/eurostat/data/database>
Rohit Singh R., Kumar S. Building Footprint Extraction and Damage Classification. [online]. <https://medium.com/geoai/building-footprint-extraction-and-damage-classification-8a5458759332>
Shekhar S., Xiong H., Zhou X. (eds) Encyclopedia of GIS. Springer, Cham. 2017, https://doi.org/10.1007/978-3-319-17885-1_432

Vedoucí diplomové práce: doc. Ing. Zdena Dobešová, Ph.D.
Katedra geoinformatiky

Datum zadání diplomové práce: 9. prosince 2022

Termín odevzdání diplomové práce: 9. května 2024

L.S.

doc. RNDr. Martin Kubala, Ph.D.
děkan



prof. RNDr. Vilém Pechanec, Ph.D.
vedoucí katedry

V Olomouci dne 1. září 2023

OBSAH

ZOZNAM POUŽITÝCH SKRATIEK	9
ÚVOD	10
1 CIELE PRÁCE.....	11
2 SÚČASNÝ STAV RIEŠENEJ PROBLEMATIKY	12
2.1 Pojmy a kľúčové slová	12
2.2 Zdroje a články	15
3 METÓDY A POSTUP SPRACOVANIA.....	22
4 ARCGIS NOTEBOOKY.....	27
4.1 Štartovací balíček.....	27
4.1.1 Základy Jupyter Notebooku	27
4.1.2 Python cvičenia	29
4.1.3 ArcPy cvičenia	31
4.2 Kolokačná analýza	34
4.3 Hustota dopravnej siete.....	43
4.4 Eurostat	48
4.5 Rozloha řadovca.....	58
4.6 Mapovacie knižnice	63
5 VÝSLEDKY	72
6 DISKUSIA.....	74
7 ZÁVER	77
POUŽITÁ LITERATÚRA A INFORMAČNÉ ZDROJE.....	78
PRÍLOHY	

ZOZNAM POUŽITÝCH SKRATIEK

Skratka	Význam
API	Application programming interface
B03	zelené pásmo B03
B11	krátkovlnné infračervené pásmo B11
CLQ	globálny kolokačný koeficient
Colab	Google Colaboratory
ČSÚ	Český statistický úřad
DPZ	dialkový prieskum Zeme
EÚ	Európska Únia
FUA	Functional Urban Area
GDP	gross domestic product – hrubý domáci produkt
GIS	geografický informačný systém
GISCO	Geographic Information System of the COMmission
GNSS	Global navigation satellite system
HRST	Human resources in science and technology
HTML	Hypertext Markup Language
ID	identifikátor
KGI	Katedra geoinformatiky PrF UP
KGI/MEDPZ	Pokročilé metody DPZ
KGI/PRODA	Programové zpracování dat
KGI/PROPY	Programování v geoinformatice
L2A	Level-2A
LCLQ	lokálny kolokačný koeficient
NDGI	Normalized Difference Glacier Index
NDSI	Normalized Difference Snow Index
NDVI	Normalized Difference Vegetation Index
NIR	Near infrared band – blízke infračervené pásmo (B8)
NUTS	Nomenclature of territorial units for statistics
OSM	OpenStreetMap
RAM	Random Access Memory
RED	Red – červené pásmo (B4)
RGB	Red Green Blue
shp	shapefile
tif	tagged image file format
UA LULC	Urban Atlas Land Use/Land Cover
URL	Uniform Resource Locator
USD	United States dollar – americký dolár
UTM	Universal Transverse Mercator
VS Code	Visual Studio Code
ZÚ	Zeměměřický úřad

ÚVOD

ArcGIS Notebook je Jupyter Notebook implementovaný do ArcGIS prostredia. Jupyter Notebook je dokument kombinujúci programový kód so sprievodným textom a s vizualizáciami dát napr. vo forme máp a grafov. Základnou charakteristikou je bunková štruktúra, kde sa striedajú bunky kódu s markdown bunkami. Táto štruktúra ponúka nový spôsob programovania, pretože bunky sa dajú spúšťať postupne a nezávisle od seba. Kód je tak rozdelený na viacero menších častí, takže vývojár môže bez problémov prototypovať, to znamená ladiť a testovať určitú funkcionality alebo algoritmus bez toho, aby musel spúšťať celý kód skriptu. Z rozdelenia kódu do buniek profituje aj užívateľ, ktorý dostane už hotový notebook, pretože získava tzv. „pohľad dovnútra“ kódu. Vďaka sprievodnému textu chápe, čo konkrétna bunka kódu robí a zároveň to môže otestovať jej spustením, po ktorom vidí výsledok priamo pod bunkou. Štýl používania a spúšťania kódu v Jupyter Notebooku sa dá nazvať interaktívne programovanie. Jupyter Notebook podporuje viacero programovacích jazykov, medzi inými aj Python, ktorý ponúka širokú škálu knižníc na prístup, spracovanie, analýzu a vizualizáciu dát a zároveň je jazykom, ktorým sa dá pristupovať k nástrojom ArcGIS Pro pomocou knižnice ArcPy.

Keďže ide o relatívnu novinku, motiváciou riešenia tejto témy je predstaviť ArcGIS Notebook a preskúmať možnosti, ktoré ponúka spracovanie a vizualizácia priestorových dát touto formou. Vedľajšou motiváciou je porovnanie notebooku a klasického Python skriptu.

Témy spracované vo výstupných notebookoch sú:

- Kolokačná analýza
- Hustota dopravnej siete
- Automatizovaný prístup k dátam z Eurostatu
- Odhad rozlohy ľadovca na základe spektrálneho indexu
- Python knižnice vytvárajúce mapy

Spoločným základom výstupných notebookov je spracovanie európskych dát. Výstupmi sú aj 3 notebooky, ktoré sú úvodom do problematiky ArcGIS/Jupyter Notebookov, obsahujú tutoriál, ako začať pracovať s notebookmi a tiež cvičenia, na ktorých je možné trénovať základné zručnosti z jazyka Python a z knižnice ArcPy.

Notebook Hustota dopravnej siete nadväzuje na skript z bakalárskej práce Možnosti výpočtu charakteristik dopravní sítě států a měst (Tóth 2022) a notebook Rozloha ľadovca je založený na skripte zo semestrálnej práce Geopython: výpočet rozlohy ľadovca (Potočiar a Tóth 2023). Obe práce boli vypracované na KGI (Katedra geoinformatiky PŕF UP). Notebook Mapovacie knižnice je inšpirovaný článkom 6 python libraries to make beautiful maps (Rozanov 2024) publikovaným na platforme Medium.

1 CIELE PRÁCE

„Cieľom diplomovej práce je vytvoriť vo formáte ArcGIS Notebook vhodný data-storytelling týkajúci sa európskych miest a štátov. Študent vyberie vhodné témy analýzy ako je doprava, využitie územia a ďalšie, ktoré formou ArcGIS Notebooku analyzuje, doplní komentármi, tabuľkami, grafmi a mapami. ArcGIS Notebooky budú slúžiť na spracovanie dát formou „dát s príbehom“ tak, aby užívateľ formou sprievodného textu a automatizovaného spúšťania čiastočných úsekov Python kódu bol komplexne zoznámený s dátami a výsledkami priestorových analýz pre vybrané témy európskych miest a štátov“ (Zadanie tejto diplomovej práce, 2022).

Vybrané témy notebookov budú zahŕňať využitie územia v mestskom prostredí, hustotu dopravnej siete, automatizovaný prístup k dátam z Eurostatu, spracovanie satelitných snímok a tvorbu máp pomocou Python knižníc. Zdrojovými dátami budú európske dáta, resp. dáta z územia európskeho svetadiela, štátov a miest.

Študent vytvorí minimálne 4 notebooky, ktoré budú spracovávať vybrané témy na uvedených dátach. Notebooky budú obsahovať sprievodný vysvetľujúci text, ktorý užívateľa uvedie do problematiky, vysvetlí programový kód a interpretuje výsledky analýz a vizualizácií. Programový kód v notebookoch bude napísaný v jazyku Python s využitím open-source aj proprietárnych (ArcPy) knižníc. Text aj kód budú napísané v anglickom jazyku. Dôležitou súčasťou každého notebooku budú vizualizácie vo forme tabuliek, grafov a máp. Výsledné notebooky môžu nadväzovať na iné bakalárske a diplomové práce, odborné články alebo existujúce ArcGIS/Jupyter Notebooky.

Ďalším cieľom tejto diplomovej práce je predstaviť ArcGIS/Jupyter Notebooky ako programovací dokument, preskúmať možnosti, ktoré ponúkajú a demonštrovať ich na príkladoch.

Užívateľ vďaka postupnému spúšťaniu čiastočných úsekov kódu získa prehľad o spracovanej téme a dátach a jednoduchšie pochopí jednotlivé algoritmy. Vytvorené notebooky tak „spopularizujú“ programovanie v Pythone a spracovanie a vizualizáciu geografických dát formou notebooku. Ich ďalším prínosom je potenciál využitia notebookov vo výukovom procese na univerzitách, kde sa študuje geoinformatika.

2 SÚČASNÝ STAV RIEŠENEJ PROBLEMATIKY

2.1 Pojmy a kľúčové slová

ArcGIS Notebook

ArcGIS Notebook je založený na Jupyter Notebooku, ktorý je implementovaný do ArcGIS prostredia. Výhodou je prepojenie robustnej ArcGIS funkcionality a množstva nástrojov geoprocessingu s Python open-source knižnicami na spracovanie dát. „Zatiaľ čo notebooky môžu byť používané a spúšťané lokálne v ArcGIS Pro, Esri integrované prostredie Jupyter Notebooku je k dispozícii aj v ArcGIS Online alebo na portáli ArcGIS Enterprise“ (Esri no date).

Jupyter Notebook

Jupyter Notebook je dokument kombinujúci programový kód s vysvetľujúcim textom, s dátami a s vizualizáciami vo forme máp, grafov a obrázkov. Má bunkovú štruktúru, pričom na výber sú 2 typy buniek: kód a markdown. Bunka je samostatné textové pole, do ktorého je možné písať a ktoré sa dá spúšťať. Notebook je jeden súbor s koncovkou .ipynb.

Jupyter Notebook je jedným z produktov projektu Jupyter, čo je „neziskový open-source projekt, ktorý vznikol z projektu IPython v roku 2014. Jeho cieľom je podporovať interaktívnu dátovú vedu a vedecké výpočty naprieč všetkými programovacími jazykmi“ (Jupyter Development Team no date). Názov Jupyter vychádza z názvov 3 programovacích jazykov, ktoré boli projektom podporované ako prvé: Julia, Python a R. V súčasnosti projekt podporuje viac ako 40 programovacích jazykov. Logo projektu (Obr. 1) symbolizuje planétu Jupiter s jeho mesiacmi.



Obr. 1 Logo projektu Jupyter (zdroj: https://en.wikipedia.org/wiki/File:Jupyter_logo.svg).

A aké sú výhody Jupyter Notebooku? Prečo na programovanie používať notebook namiesto klasického Python skriptu (súbor s koncovkou .py)? Tým, že notebook rozdeľuje kód na menšie časti prostredníctvom buniek sa programovanie stáva naozaj interaktívnym a jednoduchším na pochopenie. Užívateľ má možnosť spustiť menšiu časť kódu a výsledok vidí priamo pod spustenou bunkou, nemusí sa pozeráť do konzole. Umožňuje „rozmeniť algoritmus na drobné“ a tým nahliadnuť do jeho vnútra. Preto je notebook vhodný napríklad na výuku a na vysvetľovanie tém náročnejších na pochopenie.

Ďalšou činnosťou, na ktorú je vhodný Jupyter Notebook, je tzv. „data storytelling“. Je to efektívna komunikácia poznatkov zo skúmaného datasetu a vytvorenie určitého

príbehu, ktorý dáta rozprávajú. „Data storytelling“ tvoria 3 hlavné časti: dáta, rozprávanie (vysvetľujúci text, komentáre k poznatkom z dát) a vizualizácie (Cote 2021). Všetky 3 časti vedia byť realizované v Jupyter Notebooku s použitím programovacieho jazyka Python, ktorý vyniká v práci s dátami, obsahuje množstvo knižníc na sťahovanie, spracovanie, analýzu a vizualizáciu dát. Vizualizácie (grafy, mapy, obrázky, tabuľky) sú umiestnené pod bunkou kódu, ktorým boli vytvorené a sú tak súčasťou notebooku.

Nieto (2017) uvádza 3 dôvody, prečo by GIS (geografický informačný systém) užívatelia mali používať Jupyter Notebook:

1. Prototypovanie Python pracovných postupov – možnosť testovať určitú časť kódu bez toho, aby bol spúšťaný celý skript. Každá bunka kódu sa dá spúšťať samostatne nezávisle na ostatných kódových bunkách.
2. Vizualizácia dátových štruktúr Pandas DataFrames – DataFrame je v podstate tabuľka uložená v pamäti počítača a Jupyter Notebook vizualizuje túto tabuľku prívetivým grafickým rozhraním. Výhodou je že si užívateľ vie zobrazíť túto tabuľku na akomkoľvek mieste v notebooku a vidí, v akom stave sa nachádzajú jeho dáta.
3. Integrácia s ArcGISom – Jupyter Notebook je možné otvárať, editovať a spúšťať v ArcGIS Pro. Zároveň ArcGIS API (Application programming interface) pre Python prepája Jupyter Notebook s ArcGIS Online a ArcGIS Enterprise a umožňuje vytvárať webové mapy priamo v notebooku.

Kernel

Kernel v Jupyter Notebooku je tzv. „runtime environment“, ktoré poskytuje podporu programovaciemu jazyku, v ktorom je napísaný kód notebooku (IBM documentation 2021). Kernel je zodpovedný za spúšťanie kódu, inými slovami, je jadrom, motorom notebooku. V tejto práci je programovacím jazykom Python, takže je potrebné vybrať nejaký Python kernel. Tým môže byť stiahnutý a nainštalovaný obyčajný Python, prostredie vytvorené pomocou distribúcie Condy (Anaconda, Miniconda, a iné), alebo naklonované ArcGIS Python prostredie.

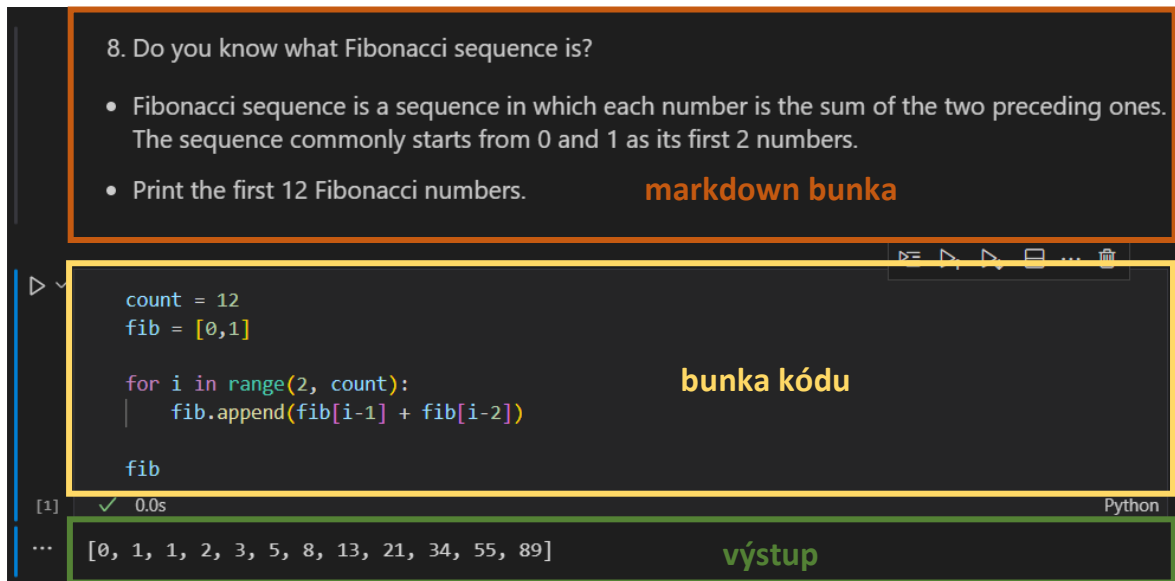
Markdown bunky

Ako už bolo spomenuté, bunky v Jupyter Notebooku sa delia na 2 základné typy: markdown a kód. Markdown bunky sú primárne určené na písanie textu, ktorý vysvetľuje kód a tému notebooku. Text je možné formátovať, takže je možné vytvoriť nadpisy, číslované a nečíslované zoznamy, hyperlinky, zvýrazniť text tučne alebo kurzívou, vkladať obrázky, LaTeX rovnice atď. „Markdown je populárny značkovací jazyk, ktorý je nadmnožinou jazyka HTML (Hypertext Markup Language)“ (Jupyter Team 2015). Markdown syntax je preto veľmi podobná HTML syntaxi.

Bunky kódu

Tieto bunky obsahujú kód vo vybranom programovacom jazyku, ktorým je v tomto prípade Python. Kód je spúšťaný kernelom, ktorý je vysvetlený vyššie v časti Kernel. Aby sa Python prostredie mohlo stať kernelom Jupyter Notebooku, musí mať nainštalovaný balík ipykernel. Bez neho totiž nevie spúšťať kód v bunkách notebooku. Existuje viacero spôsobov, ako inštalovať balíky a knižnice do Python prostredia, jedným z nich je napísať príkaz na inštaláciu do terminálu priamo vo VS Code (Visual Studio Code). Viac o inštalácii knižníc do ArcGIS Python prostredia je napísané v podkapitole 4.1.3 pri notebooku ArcPy cvičenia, ktorý sa priamo prakticky venuje úvodu práce s Jupyter Notebookom v ArcGIS

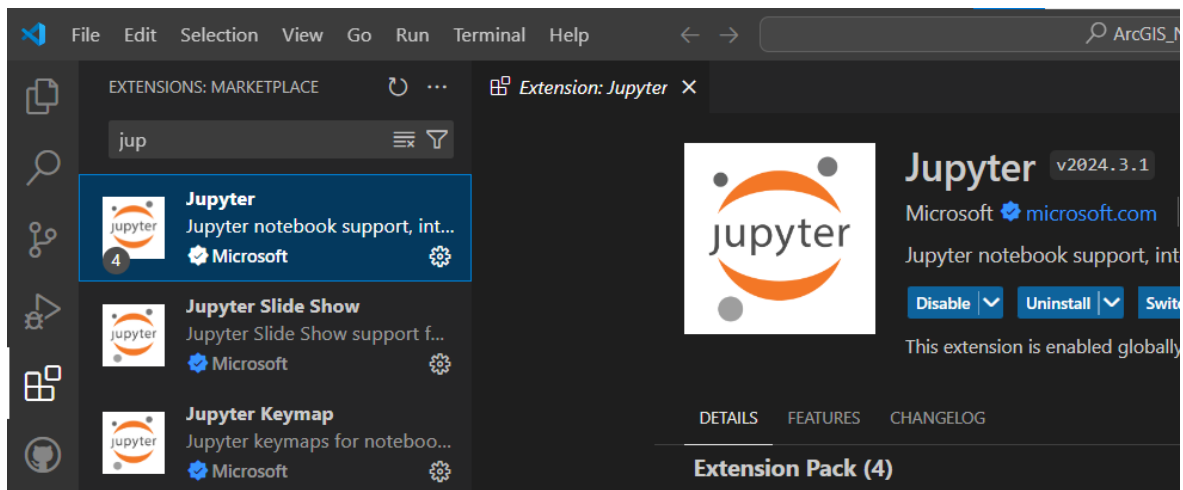
Pro. Bunková štruktúra notebooku založená na striedaní markdown buniek a buniek kódu je ukázaná na Obr. 2.



Obr. 2 Ukážka markdown bunky a spustenej bunky kódu s výstupom (autor: A. Tóth).

Visual Studio Code

VS Code je v súčasnosti jeden z najpopulárnejších a najrozšírenejších editorov programového kódu a poskytuje vhodné užívateľské rozhranie na prácu s Jupyter Notebookmi. Ďalšou výhodou je to, že je šírený pod open-source MIT licenciou, čo umožňuje jeho voľné stiahnutie. Okrem toho podporuje zálohovací a verzovací systém Git a farebne zvýrazňuje kód, dopĺňa názvy premenných a funkcií a kontroluje syntax kódu. Jedinou prerekvizitou na používanie VS Code na tvorbu a spúšťanie Jupyter Notebookov je inštalácia extenzie Jupyter (Obr. 3). Táto inštalácia je zadarmo a je rýchla.



Obr. 3 Extenzia Jupyter pre VS Code (autor: A. Tóth).

Google Colaboratory a JupyterLab

Okrem VS Code existujú samozrejme aj ďalšie prostredia, v ktorých sa dajú vytvárať, editovať a spúšťať Jupyter Notebooky. Medzi najznámejšie patrí [Colab](#) (Google

Colaboratory), čo je hostovaná služba pre Jupyter Notebooky poskytujúca prístup k výpočtovým kapacitám vrátane GPU a TPU (Google no date). Colab je jednou z mnohých Google „extenzií“ alebo „aplikácií“, ktoré sú k dispozícii tým, ktorý majú vytvorený Google účet. Je to možné porovnať k Esri aplikáciám ako Field Maps, Quick Capture, Survey 123, Dashboards atď.: všetky zdieľajú jedno cloudové úložisko v rámci ArcGIS Online. Rovnako aj Google Sheets, Documents, Colab a ďalšie aplikácie zdieľajú úložisko Google Drive. Do Colab je možné nahráť Jupyter Notebooky z lokálneho úložiska na počítači, z GitHubu alebo Google Disku. Kernel je v Colab pomenovaný ako „runtime environment“ a je možné pripojiť sa k takémuto hostovanému prostrediu s GPU. Ďalšie možnosti sú pripojiť sa k vlastnému virtuálnemu počítaču GCE alebo k miestnemu prostrediu runtime.

„[JupyterLab](#) je autorská, vysoko rozšíriteľná a funkčne bohatá aplikácia na tvorbu a editáciu notebookov“ (Jupyter Team 2024). Je to pôvodné webové užívateľské rozhranie projektu Jupyter, ktoré vie pracovať s notebookmi, textovými editormi a terminálmi. Umožňuje pripojiť sa k viacerým kernelom, Python kernel má názov Pyodide.

2.2 Zdroje a články

Ukážkové Arcgis notebooky

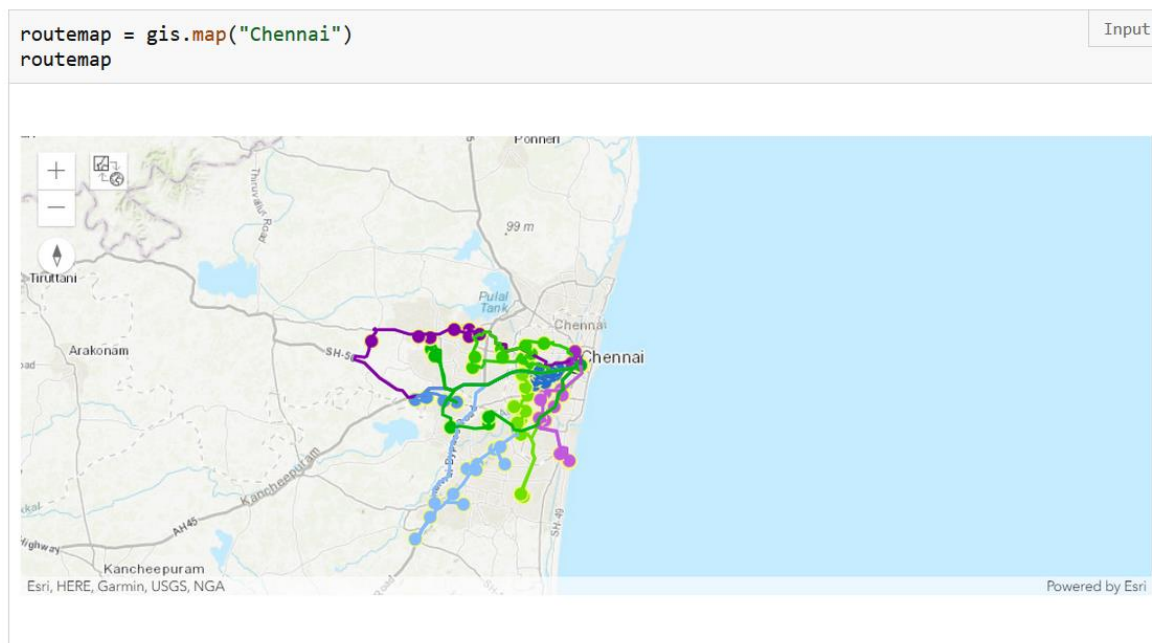
ArcGIS API pre Python má na svojej [webovej stránke](#) množstvo ukážkových notebookov, ktoré demonštrujú rôzne funkcie a možnosti tohto API. Sú rozdelené do 5 kategórií podľa cieľových skupín: vývojári, administrátori, GIS analytici a dátoví vedci, hlboké učenie, publikovanie obsahu. V kategórii pre GIS analytikov a dátových vedcov sa nachádzajú 2 notebooky, ktoré boli prvými inšpiráciami pre túto prácu.

Tým prvým je notebook s názvom **Chennai Floods 2015 – A Geographic Analysis**. Notebook mapuje rozsah, príčiny a dôsledky povodní, ktoré zasiahli oblasť okolo indického mesta Chennai v na konci roku 2015. „Počas 1. a 2. decembra 2015 spadlo v Chennai najviac dažďových zrážok za 24 h od roku 1901. Najmenej 250 ľudí zomrelo, niekoľko stoviek ľudí bolo kriticky zranených a tisícky boli priamo ovplyvnené alebo sa museli premiestniť v dôsledku záplav“ (Esri - ArcGIS API for Python no date). Notebook je rozdelený na viacero častí, v ktorých používa rôzne dátové zdroje z ArcGIS Online ako napr. bodovú vrstvu centier pomoci, Landsat GLS multispektrálne snímky, polygónovú vrstvu jazier v okolí Chennai a bodovú vrstvu meteorologických staníc s úhrnom zrážok. Tieto priestorové dáta sú vstupom pre ďalšie spracovanie, analýzy a vizualizácie. Na tvorbu grafov je použitá knižnica Matplotlib a na vizualizáciu priestorových dát je využívané rozhranie ArcGIS Online, cez ktoré je možné vytvárať interaktívne webové mapy priamo v notebooku.

Prvé dve časti sú zamerané na zmapovanie rozsahu privalových dažďov. Na vytvorenie spojitého povrchu reprezentujúceho úhrn zrážok v skúmanej oblasti bola použitá interpolácia bodovej vrstvy meteorologických staníc pomocou empirického Bayesovského krigingu. Záplavy boli okrem samotných zrážok spôsobené aj vyliatím riek, jazier a vodných nádrží, ktoré boli v dôsledku dažďov preplnené. V ďalšej časti notebooku sú preto pomocou trasovania po prúde a vytvoreniu bufferov určené oblasti, ktoré mohli byť najviac zasiahnuté vyliatím vodných útvarov v krajine. Autor notebooku sa zaoberá otázkou, či boli záplavy spôsobené prírodným živlom a klimatickou zmenou alebo skôr nesprávnym územným plánovaním a ničením prirodzených ekosystémov, ktoré vedú absorbovať veľké množstvo vody. Sústredí sa na oblasť močiarov Pallikarandai, ktoré boli kedysi miestom pre sťahovavé vtáky, no dnes sa tam nachádza skládka odpadu a mení charakter močiarov

(Madonna a Satyarupa 2015). Pomocou satelitných snímok z rokov 1991 a 2009 pozoruje, ako sa zmenilo toto prostredie v dôsledku ľudských aktivít. Na záver sú zmapované centrá pomoci, takže užívateľ má prehľad, kde sa nachádzajú a aké majú kapacity. Funkciou `plan_routes` z ArcGIS API pre Python sú vytvorené optimálne trasy pre šoférov vozidiel, ktorých úlohou je voziť humanitárnu pomoc do týchto centier (Obr. 4).

Notebook *Chennai Floods 2015 – A Geographic Analysis* je ukážkou toho, ako dokáže notebook obsiahnuť celistvú GIS analýzu udalosti, akou je prírodná katastrofa. Zároveň ide o príklad „data storytelling“, kde je pomocou dát, ich analýz, vizualizácií a naratívneho textu prerozprávajú príbeh udalosti záplav v Indii, ich príčin, dôsledkov a riešení.



Obr. 4 Mapa optimálnych trás na doručenie humanitárnej pomoci (zdroj: <https://developers.arcgis.com/python/samples/chennai-floods-analysis-rn/>).

Towards Data Science a Medium

[Towards Data Science](#) Inc je kanadská spoločnosť, ktorá prostredníctvom portálu [Medium](#) vytvára platformu pre články o dátovej vede (TDS Editors 2018). Medium je podporované komunitou jej členov, keďže niektoré články sú voľne dostupné, iné sú dostupné po vytvorení bezplatného účtu a ďalšie sú platené. Obe platformy ponúkajú množstvo článkov na témy GIS, DPZ (diaľkový prieskum Zeme), Python, Jupyter Notebook, preto boli dôležitým zdrojom pri vypracovaní tejto práce.

Článok **Using Eurostat statistical data on Europe with Python** (van der Meulen 2022) sa venuje práci s dátami z Eurostatu v Pythone. Autor článku vyvinul vlastnú knižnicu na sťahovanie Eurostat datasetov s názvom EasyEuroStat, ktorá je v článku prezentovaná. Dataset o nezamestnanosti v EÚ (Európska Únia) s kódom *teilm020* je stiahnutý ako súbor `.tsv` a následne je tento súbor načítaný do štruktúry Pandas DataFrame. Autor demonštruje skúsenosti s manipulovaním dát v štruktúrach slovník (dictionary) a DataFrame. Pomocou knižnice Matplotlib vytvára stĺpcový graf vývoja nezamestnanosti v Rakúsku počas 12 mesiacov od októbra 2021 do septembra 2022. Ďalšou vizualizáciou je pseudokartogram nezamestnanosti v EÚ v auguste 2022. Na vytvorenie mapy bolo potrebné stiahnuť priestorovú vrstvu krajín EÚ vo formáte shp

(shapefile) z rozhrania GISCO (Geographic Information System of the COMmission) services. Po stiahnutí bola táto vrstva načítaná do štruktúry GeoDataFrame a boli k nej pripojené dáta o nezamestnanosti. Výsledná mapa vo forme statického obrázka bola vytvorená funkciou plot, ktorá zo štruktúry GeoDataFrame vytvorí mapu založenú na knižnici Matplotlib.

Stahovanie Eurostat datasetov na základe ich ID (identifikátor) má výhodu v tom, že zakaždým sú stiahnuté aktuálne dáta. To zvyšuje význam nasadenia notebooku, kedy je možné pomocou existujúceho notebooku rýchlo vytvoriť aktuálne výstupy. Tento článok bol jednou z inšpirácií pre notebook Eurostat, ktorý je jedným z výstupov tejto diplomovej práce.

Článok **Plotting maps with European data in Python: Part I** (Hapala 2021) sa zameriava na vizualizáciu vrstvy európskych štátov pomocou knižníc GeoPandas a Contextily. Autor ako zdroj dát tiež využíva GISCO, avšak nestahuje ich priamo v kóde ale „ručne“ a prikladá hyperlink na webovú stránku. Pracuje s vrstvami európskych krajín, ich hraníc a popiskov, ktoré sťahuje vo formáte TopoJSON v súradnicovom systéme ETRS 1989 LAEA (epsg: 3035). Stiahnuté vrstvy sú načítané do samostatných štruktúr GeoDataFrame a jednoduchá mapa je vytvorená funkciou plot.

Nasleduje časť, ktorá vysvetľuje princíp súradnicových systémov a kartografických zobrazení a porovnáva 3 vybrané projekcie: ETRS 1989 LAEA (epsg: 3035), WGS 1984 (epsg: 4326), WGS 84/Pseudo-Mercator (epsg: 3857). Na vizualizáciu európskych krajín vo všetkých 3 zobrazeniach (Obr. 5) je potrebné uložiť vrstvu v príslušných súradnicových systémoch. Preprojektovanie štruktúry GeoDataFrame je dostupné vo funkcii *to_crs*. Pre Európu je najlepšia projekcia ETRS 1989 LAEA (epsg: 3035), ktorá síce skresľuje uhly, ale zachováva plochy.

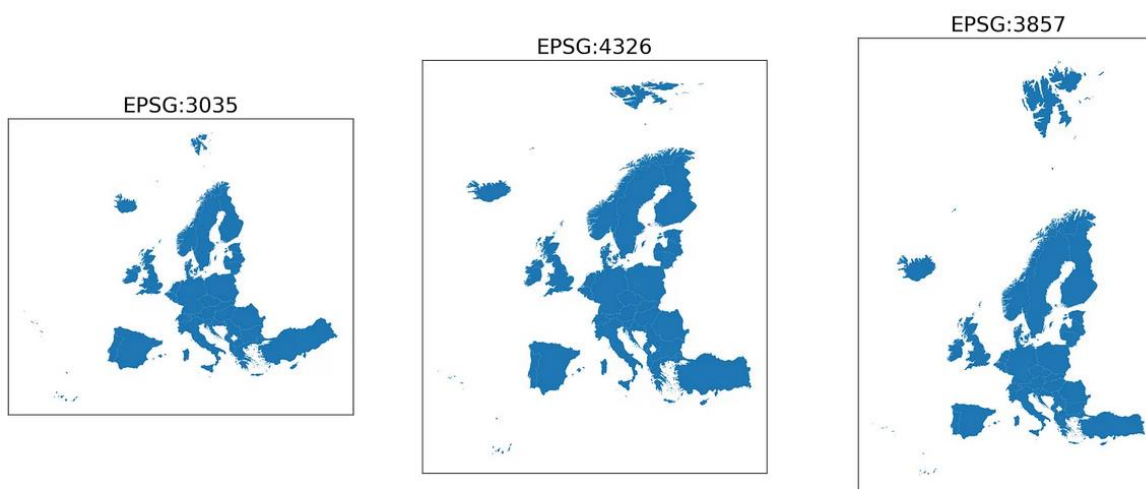
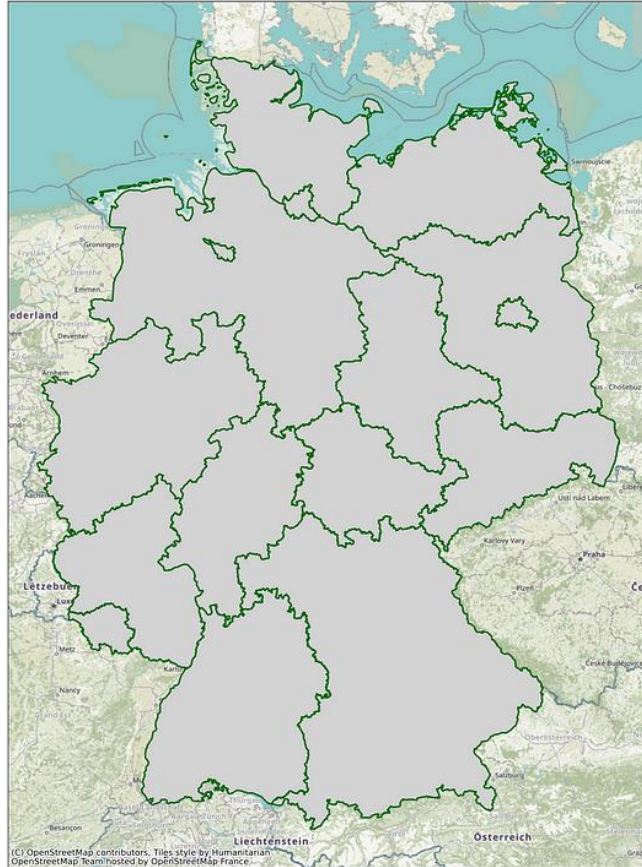


Fig. 2: Comparison of three Coordinate Reference Systems. The middle plot shows the CRS working with sphere coordinates (degrees), the other two display geometries projected on a plane. France is removed from the data to cut off overseas territories from the picture. Source of data © EuroGeographics for the administrative boundaries

Obr. 5 Porovnanie 3 projekcií na mapách zobrazujúcich európske štáty (autor: Hapala, 2021).

Na záver sú vizualizované nemecké regióny NUTS (Nomenclature of territorial units for statistics) 1, čiže nemecké spolkové republiky. Tentokrát je do mapy pridaná aj podkladová mapa OSM (OpenStreetMap) pomocou knižnice Contextily (Obr. 6).



Obr. 6 Mapa nemeckých spolkových republík s podkladovou mapou OSM (autor: Hapala, 2021).

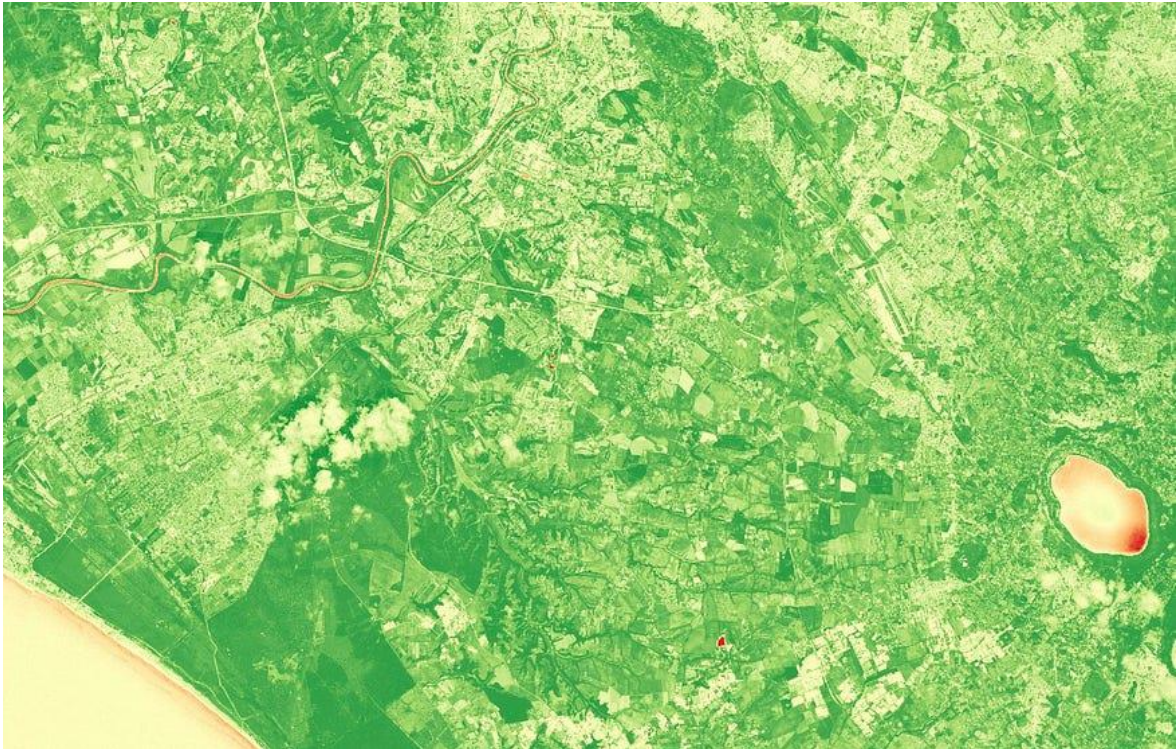
Článok **Satellite imagery access and analysis in Python & Jupyter notebooks** (Hassan 2019) je vo svojej podstate tutoriálom, ktorý sa je rozdelený na 3 časti:

- vyhľadávanie a sťahovanie satelitných snímok pomocou Python knižnice Sentinelsat priamo v Jupyter Notebooku
- čítanie a zápis rastrových dát
- vytvorenie RGB (Red Green Blue) a NDVI (Normalized Difference Vegetation Index) z pásiem Sentinel 2

Zaujímavým územím je prírodná rezervácia Castel Porziano južne od Ríma. Autor používa knižnicu Folium na vizualizáciu polygónu ohraničujúceho túto prírodnú rezerváciu. Polygón musí byť uložený vo formáte Shapely geometry, aby mohol slúžiť ako priestorový filter na získavanie satelitných snímok a následne na ich orezanie. Dotaz na získanie satelitných snímok obsahuje tieto podmienky: snímky sú z obdobia od 1.-26.6.2019, sú z misie Sentinel-2, úroveň spracovania je L2A (Level-2A) a podiel oblačnosti na snímke je medzi 0-10%. Výsledky dotazu sú uložené do štruktúry GeoDataFrame a snímka s najmenším podielom oblačnosti je stiahnutá do lokálneho úložiska počítača. Funkciami z knižnice Rasterio je zostavený RGB obrázok z pásem viditeľnej časti spektra stiahnutej snímky. Rasterio ponúka aj funkciu mask, ktorou je orezaná snímka podľa polygónu prírodnej rezervácie. Na záver je vypočítaný NDVI podľa vzorca:

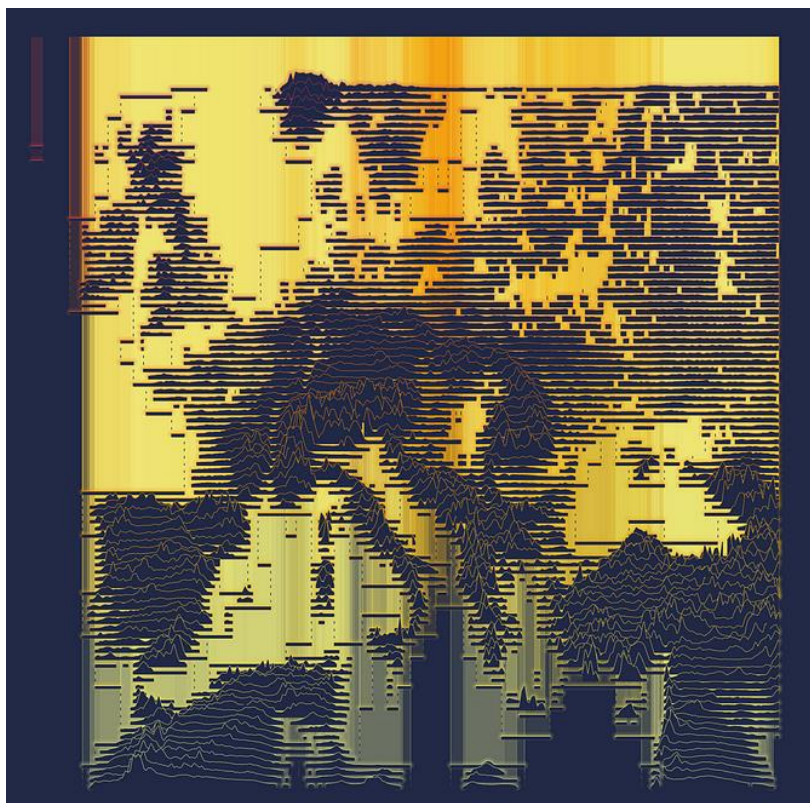
$$\text{NDVI} = (\text{NIR} - \text{RED}) / (\text{NIR} + \text{RED}),$$

kde NIR (near infrared band) je blízke infračervené pásmo (B8) a RED je červené pásmo (B4). NDVI je zapísaný ako raster do formátu tif (tagged image file format) a autor prikladá jeho vizualizáciu v programe QGIS (Obr. 7).



Obr. 7 NDVI index vizualizovaný v programe QGIS (autor: A. Hassan).

Článok **6 python libraries to make beautiful maps** (Rožanov 2024) predstavuje 6 Python knižníc, ktoré vedia vytvárať mapy. Tento článok je priamou inšpiráciou vlastného riešenia notebooku Mapovacie knižnice. Knižnica Cartopy vytvára statické mapy založené na knižnici Matplotlib. Výhodou je pestrá paleta kartografických zobrazení a zabudovaných podkladových vrstiev. Knižnica Folium je jedna z najpopulárnejších a je zameraná na interaktívne mapy postavené na Leaflet.js. Plotly je firma, ktorá okrem iného zastrešuje open-source Python knižnicu, ktorá vie produkovať interaktívne grafy a mapy postavené na Matplotlib jadre. Ďalšou výhodou je tiež možnosť vybrať si z viacerých projekcií. Knižnica ipyleaflet rovnako ako Folium používa Leaflet.js na tvorbu interaktívnych máp. Ponúka množstvo podkladových máp a zaujímavé animované vizualizácie. Geemap je knižnica určená na tvorbu interaktívnych máp s Google Earth Engine. Ponúka široké spektrum funkcií, nevýhodou je však nutnosť znalosti Google Earth Engine a fakt, že nie je užívateľsky príjemná/prívetivá. Poslednou knižnicou prezentovanou v tomto článku je ridgemap, ktorá ponúka umelecké formy vizualizácie. Ukážkou je napr. vizualizácia výškopisu v Európe (Obr. 8).



Obr. 8 Vizualizácia výškopisu v Európe knižnicou ridgemap (autor: A. Rozanov).

Esri tutoriál

Esri na svojich webových stránkach ponúka množstvo kurzov a tutoriálov. Jedným z nich je tutoriál s názvom [Naučte sa Python s ArcGIS Notebookmi \(Learn Python with ArcGIS Notebooks\)](#). Obsahuje 10 notebookov, ktoré učia užívateľov, „ako vytvoriť Python kód pre vizualizáciu, analýzu a správu dát na akejkoľvek platforme cez ArcGIS Online“ (Esri no date). Na používanie a vytváranie notebookov v ArcGIS Online je však potrebné mať extenziu licencie, ktorá nie je zahrnutá v univerzitnom ArcGIS účte poskytnutom autorovi tejto diplomovej práce. Notebooky sa však dajú stiahnuť ako súbory .ipynb a dajú sa používať lokálne, napr. v ArcGIS Pro alebo vo VS Code. Z 10 tutoriálových notebookov sa autor zameril na jeden, konkrétne notebook číslo **5. Using arcpy in ArcGIS Notebooks**. Notebook vysvetľuje, ako používať knižnicu ArcPy a kedy použiť ArcGIS API pre Python namiesto ArcPy. Obsahuje návod, ako nahrať súborovú geodatabázu do prostredia ArcGIS Online a ako vizualizovať dáta z tejto geodatabázy na mape priamo v notebooku. Tento tutoriálový notebook je prehľadne členený do niekoľkých častí, ktorých hyperlinkový zoznam sa nachádza hneď v úvode. V prvej časti sa nachádza bunka kódu, ktorý po spustení skryje všetky bunky kódu a užívateľ tak môže písať kód samostatne na základe zadania v markdown bunkách. Ak si nie je istý alebo potrebuje pomoc, klikne na odkaz, ktorý sa vygeneroval po spustení bunky so skrývajúcim kódom, a bunky kódu sa opäť ukážu. Text v markdown bunkách je užívateľsky prívetivo formátovaný (ukážka na Obr. 9), slová sú podľa potreby zvýraznené tučne alebo kurzívou, názvy premenných a Python knižníc sú zvýraznené štýlom typickým pre „citovanie“ kódu. Prínos tohto notebooku pre túto prácu spočíval predovšetkým vo formátovaní textu markdown buniek.

Create a new file geodatabase

You can also create file geodatabases and feature classes in ArcGIS Notebooks.

1. Create a file geodatabase named **Results.gdb** in the `lesson_dir` folder using `arcpy.management.CreateFileGDB()`. To do this, create a variable named `gp_res` and set it equal to `arcpy.management.CreateFileGDB()`. Run the cell.

```
gp_res = arcpy.management.CreateFileGDB(lesson_dir, 'Results.gdb')
```

Python

Obr. 9 Ukážka formátovania textu v markdown bunke v tutoriálovom notebooku (autor: A. Tóth, zdroj: [5. Using arcpy in ArcGIS Notebooks](#)).

Diplomové práce

Diplomová práca B. Šrama (2023) s názvom „Automation of processing GNSS (Global navigation satellite system) track records for designing intensity maps“, ktorá bola spracovaná na KGI, sa zaoberá problematikou vizualizácie GNSS záznamov na mapách. V rámci práce je vyvinutý nástroj, ktorý „prichytáva“ GNSS záznamy na cestnú sieť. Tento nástroj je spracovaný vo forme Jupyter Notebooku.

Diplomová práca „Automatizace zpracování geografických dat pomocí ArcGIS Pro a programovacího jazyka Python“, ktorá bola spracovaná na Provozně ekonomické fakultě České zemědělské univerzity v Prahe, rieši „celý proces spracovania dát od ich počítačového stiahnutia zo vzdialeného serveru až po publikáciu v ArcGIS Online“ (Černý 2024). Tento proces je zautomatizovaný v skripte, ktorý je výstupom práce. Autor sa v značnej miere venuje aj ArcGIS Notebooku, predstavuje ho v rešerši, uvádza jeho výhody.

3 METÓDY A POSTUP SPRACOVANIA

Použité metódy

Na začiatku práce bolo potrebné zoznámiť sa s témou a zistiť, čo je možné robiť v ArcGIS/Jupyter Notebookoch. V tejto fáze boli užitočné ukážkové ArcGIS Notebooky, ako napr. notebook „Chennai Floods 2015 – A Geographic Analysis“ (Esri - ArcGIS API for Python no date) a oficiálna stránka projektu Jupyter (Jupyter Development Team no date). S vedúcou práce boli vybrané témy jednotlivých notebookov tak, aby bol každý zameraný na niečo iné a za jednotiaci prvok boli zvolené európske dáta.

Nasledovalo čítanie ďalších zdrojov a článkov napr. „Using Eurostat statistical data on Europe with Python“ (van der Meulen 2022) a popri tom boli programované prvé riadky notebookov. Samotné programovanie spočívalo v skúšaní jednoduchších častí kódu, na ktoré sa nabaľovali zložitejšie a väčšie časti, ako for cykly a opakovateľne použiteľné funkcie. Inšpirácia pri programovaní bola čerpaná z dokumentácií príslušných knižníc, kde bola zároveň overovaná správnosť syntaxe. Dôležité rozhodnutia, ktorým smerom posúvať a vyvíjať notebooky, boli konzultované s vedúcou práce.

Do naklonovaného ArcGIS Python prostredia, ktoré slúžilo ako kernel a spúšťalo kód notebookov, boli potrebné knižnice inštalované príkazom `pip install [názov knižnice]` v Python príkazovom riadku (Python command prompt). Týmto príkazovým riadkom je možné pristupovať k aktívnemu ArcGIS Python prostrediu a inštalovať doňho knižnice.

Kľúčové zdroje, z ktorých riešenie cieľov práce vychádza sú tieto:

- How Colocation Analysis works (Esri no date)
- Skript [eu_grid_population.py](#) z bakalárskej práce Možnosti výpočtu charakteristik dopravní sítě států a měst (Tóth 2022)
- Eurostat Python Package (Cazzaniga 2024)
- Dataset Human resources in science and technology (HRST) by NUTS 2 regions (Eurostat 2024)
- Semestrálna práca GeoPython: výpočet rozlohy ľadovca (Potočiar a Tóth 2023)
- Článok 6 python libraries to make beautiful maps (Rožanov 2024)

Testovanie notebookov vo výuke prebiehalo na cvičeniach predmetov KGI/PROPY (Programování v geoinformatice) v rokoch 2023, 2024 a KGI/PRODA (Programové zpracování dat) v roku 2024 (Obr. 10).



Obr. 10 Fotografia počas výuky s ArcGIS Notebookmi v rámci predmetu KGI/PRODA (autor: M. Vojtěchovská).

Študenti dostali Jupyter Notebooky, ktoré boli zároveň prezentáciou (obsahovali teoretické základy tvorby notebookov a jazyka Python) a praktickým cvičením, kedy podľa zadania v markdown bunke mali sami napísať programový kód. Najskôr bol vždy ponechaný čas, aby študenti mohli sami rozmýšľať a prísť na riešenie. To bolo vysvetlené všetkým vo chvíli, keď už naňho študenti prišli alebo kedy naopak nevedeli, ako to vyriešiť. Vždy bola aplikovaná snaha naviesť na správne riešenie, než aby bol správny kód iba jednoducho prezradený. Na konci cvičenia bola prebratá téma zhrnutá v kvíze kahoot.

Skrývanie kódu v Jupyter Notebookoch je možné a hodí sa práve na výuku. V ArcGIS Pro je to jednoducho realizovateľné pomocou extenzie `hide-input-all` alebo pomocou špeciálneho HTML kódu z knižnice `IPython.display`. Vo VS Code je skrývanie kódu oveľa náročnejšie.

Použité dáta

V jednotlivých notebookoch boli použité nasledujúce dáta:

- ArcČR 500 verzia 3.3 „je digitálna vektorová geografická databáza Českej republiky spracovaná na úrovni podrobnosti 1 : 500 000“ (ARCDATA PRAHA et al. 2016). Údaje pre vznik tejto dátovej sady boli poskytnuté od ZÚ (Zeměměřický úřad) a ČSÚ (Český statistický úřad). Je rozdelená na 2 súborové geodatabázy: `AdministrativniCleneni_v13.gdb` a `ArcCR500_v33.gdb`. Z prvej uvedenej geodatabázy bola použitá bodová a polygónová vrstva obcí, z druhej geodatabázy boli použité vrstvy s názvami `Lesy`, `SidlaPlochy`, `Silnice_2016`, `VodniToky`. Tieto vrstvy boli vstupnými dátami v notebooku ArcPy cvičenia (podkapitola 4.1.3). Dátová sada ArcČR 500 verzia 3.3 bola získaná v rámci výuky v prvom ročníku bakalárskeho štúdia odboru Geoinformatika a geografie na KGI.

- Urban Atlas Land Cover/Land Use 2018 „poskytuje dáta o pokryve a využití krajiny spolu s odhadmi populácie pre 788 európskych mestských oblastí – FUA (Functional Urban Area), ktoré majú viac ako 50 000 obyvateľov“ (Copernicus Land Monitoring Service 2020). Dáta pre konurbáciu Katowice v Poľsku boli stiahnuté po bezplatnom vytvorení účtu na webovej stránke [Copernicus Land Monitoring Service](#). Stiahnuté dáta sú uložené vo formáte geopackage a okrem toho majú priloženú aj dokumentáciu v pdf a symbológiu, resp. štýl zobrazenia vo formátoch .lyr, .sld a .qml. Sú vstupom v notebooku Kolokačná analýza (podkapitola 4.2) a keďže sú šírené ďalej spolu s týmto notebookom, je potrebné ich citovať týmto spôsobom: „European Union's Copernicus Land Monitoring Service information; DOI: <https://doi.org/10.2909/fb4dffa1-6ceb-4cc0-8372-1ed354c285e6>“.
- Železničná sieť mesta Liverpool z OSM je vstupom pre kód v notebooku Hustota dopravnej siete (podkapitola 4.3). Je stiahnutá pomocou Python knižnice OSMnx do štruktúry GeoDataFrame. OSM sú otvorené geografické dáta vytvárané komunitou prispievateľov, ktoré pod licenciou Open Database License poskytuje nadácia OpenStreetMap Foundation (OpenStreetMap contributors no date).
- GEOSTAT 2018 1 km² populačný grid je zdrojom dát o európskej populácii. Je to štvorcová mriežka o veľkosti štvorca 1 × 1 km pokrývajúca osídlené územia v Európe a každý štvorec obsahuje údaj o počte obyvateľov, ktorý tam žijú. Tento dataset vznikol „kombináciou dát o počte obyvateľov v roku 2018 poskytnutých národnými štatistickými úradmi, GEOSTAT populačného gridu z roku 2011 a dát o hustote zastavaných plôch v rokoch 2012 a 2018 odvodených z produktov Copernicus Earth Observation“ (Batista e Silva et al. 2021). Je vstupom pre notebook Hustota dopravnej siete (podkapitola 4.3), kde je stiahnutý pomocou knižnice Requests z webovej stránky [GISCO](#). Stiahnutý produkt je v rastrovom formáte tif a vo vektorovom formáte shp.
- Eurostat dataset Human resources in science and technology (HRST) by NUTS 2 regions je hlavným vstupom a témou notebooku Eurostat (podkapitola 4.4). Pomocou knižnice Eurostat je uložený v štruktúre DataFrame. Tento dataset obsahuje percentuálne podiely aktívnej populácie vo veku 15-74 rokov, ktorá buď úspešne dokončila terciárne vzdelanie alebo je zamestnaná vo vede a v technológii pre regióny NUTS 2 (Eurostat 2024).
- Polygónové vrstvy európskych regiónov NUTS 2 a 0 stiahnuté pomocou knižnice Requests z rozhrania [GISCO services](#) sú vstupnými dátami pre notebooky Eurostat a Mapovacie knižnice (podkapitoly 4.4 a 4.6). Obe vrstvy sú sťahované vo formáte GeoJSON, v súradnicovom systéme WGS 1984 (espg: 4326) a na úrovni podrobnosti 1 : 1 000 000. Povolenie používať tieto dáta je možné za podmienok, že to nie je na komerčné účely a že bude uvedený zdroj formou tohto upozornenia na autorské práva: © EuroGeographics pre administratívne hranice (Eurostat no date).
- Satelitné snímky Sentinel-2 L2A stiahnuté v júlových termínoch v rokoch 2018-2023 pre oblasť ľadovca Aletschgletscher vo švajčiarskych Alpách sú vstupnými dátami pre notebook Rozloha ľadovca (podkapitola 4.5). Po bezplatnej registrácii sú dostupné na [Copernicus Browser](#), odkiaľ boli stiahnuté aj v prípade tejto práce. Spomínaný notebook potrebuje z týchto snímok len pásma B03 a B11, čo je zelené a krátkovlnné infračervené pásmo. Tieto pásma sú uložené v zložke data v adresári, kde sa nachádza aj notebook, a preto je potrebné ich citovať: „Copernicus Sentinel data 2018-2023“ (European Commission 2017).

Použité programy

Na vývoj notebookov, to znamená na programovanie a na písanie sprievodného textu v markdown bunkách bol použitý program Visual Studio Code verzia 1.88. Notebooky, ktoré nepotrebujú ArcGIS Pro, môžu byť používané a spúšťané priamo vo VS Code.

Na zálohovanie a verzovanie notebookov bol použitý Git a GitHub. VS Code podporuje zálohovanie týmto systémom. Inšpiráciou a návodom na používanie Gitu a GitHubu boli prednášky a cvičenia z predmetu KGI/PRODA vedené Mgr. Michaelou Vojtěchovskou.

Testovanie a spúšťanie notebookov prebiehalo v programe ArcGIS Pro verzia 3.0.2. Licenciu poskytla KGI.

Programovacím jazykom bol Python verzia 3.9.18, konkrétne naklonované ArcGIS Python prostredie. Toto prostredie slúžilo ako kernel, ktorý spúšťal kód v notebookoch.

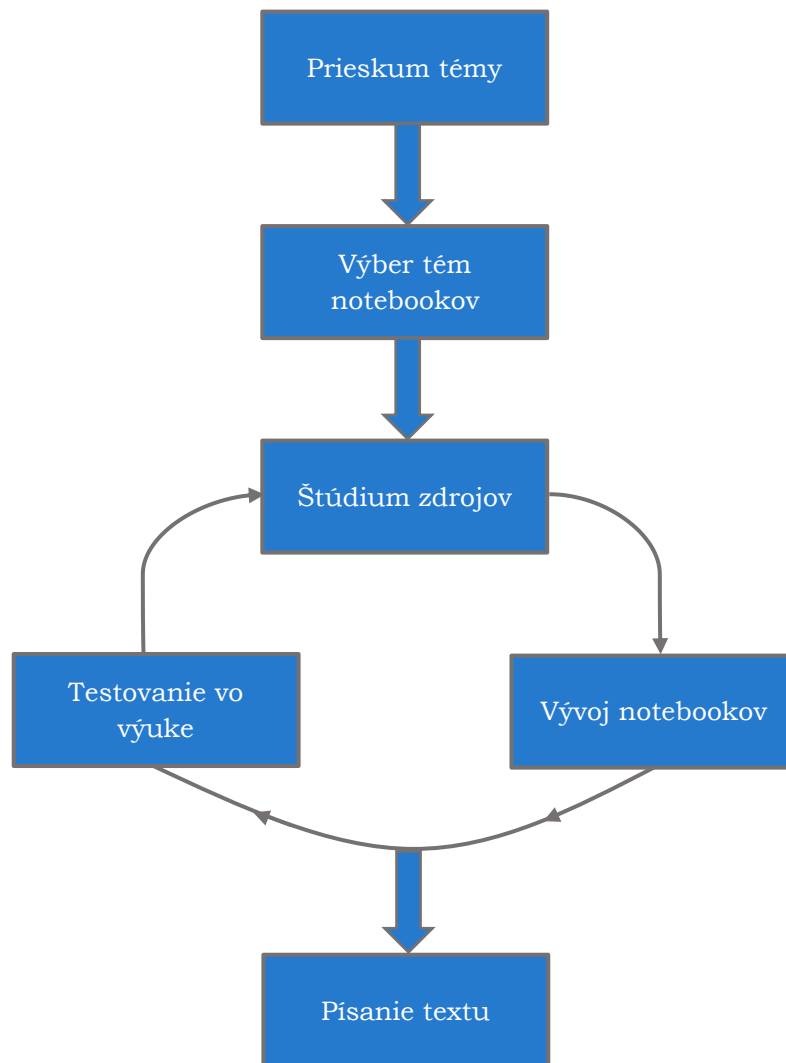
Do tohto prostredia boli inštalované potrebné Python knižnice. V tabuľke 1 je prehľad doinštalovaných knižníc a ich verzií.

Tabuľka 1 Prehľad doinštalovaných knižníc a ich verzií

Názov knižnice	verzia
Cartopy	0.22.0
Eurostat	1.1.0
Folium	0.14.0
GeoPandas	0.14.3
ipyleaflet	0.18.2
ipywidgets	7.6.5
Matplotlib	3.8.4
matplotlib-scalebar	0.8.1
Numpy	1.26.3
OSMnx	1.9.1
Pandas	2.2.1
Plotly	5.19.0
Rasterio	1.3.9
Requests	2.31.0
Shapely	2.0.3
Xarray	2022.11.0

Postup spracovania

Diagram postupu spracovania na Obr. 11 schematicky znázorňuje návaznosť jednotlivých celkov spracovania tejto práce. Prostredná cyklická časť sa opakovala niekoľkokrát, jej znázornenie nie je dokonalé, ale najlepšie vystihuje charakter práce. Štúdium zdrojov a vývoj notebookov prebiehali často súčasne a nie všetky notebooky boli otestované vo výukovom procese, ale vo všeobecnosti vytvárajú vývojový cyklus. Téma posledného notebooku o mapovacích knižniciach bola vybraná v čase, keď už ostatné notebooky boli dokončené.



Obr. 11 Diagram postupu spracovania (autor: A. Tóth).

4 ARCGIS NOTEBOOKY

V tejto kapitole sú podrobne opísané ArcGIS Notebooky, ktoré sú výstupom celej práce. Všetky notebooky sú napísané v anglickom jazyku.

4.1 Štartovací balíček

Štartovací balíček tvoria 3 notebooky, ktoré vovedú užívateľa do sveta základov jazyka Python, tvorby notebookov a ArcGIS Notebookov. Sú umiestnené v zložke /0_basics/. Všetky tri boli nasadené a testované vo výuke na predmete KGI/PROPY v letnom semestri v rokoch 2023 a 2024 a na predmete KGI/PRODA v letnom semestri v roku 2024.

4.1.1 Základy Jupyter Notebooku

Názov súboru: **0_jupyter_notebook_basics.ipynb**

Potrebný komerčný software: nie

Počet markdown buniek: 50

Počet buniek kódu: 18

Počet riadkov kódu: 40

Tento notebook je úvodom do Jupyter Notebooku, vysvetľuje jeho podstatu, význam a spôsob, ako ho používať. Jupyter Notebook je možné otvárať, používať a spúšťať jeho kód vo viacerých programoch a prostrediach, ako napr. ArcGIS Pro, JupyterLab, Google Colab, alebo VS Code. Tento notebook je určený na používanie vo VS Code, obsahuje tiež návody a popis rozhrania tohto programu. Jedinou prerekvizitou tohto notebooku je teda [stiahnutie](#) a inštalácia programu VS Code.

Prvá časť notebooku obsahuje stručné zhrnutie, čo vlastne je Jupyter Notebook, z čoho vychádza a aké sú jeho výhody. Táto téma rovnako ako kernel, markdown bunky a bunky kódu je viac rozpísaná v rešerši tejto práce v 2. kapitole.

V ďalšej časti sú pomocou snímkov obrazovky popísané dva spôsoby, ako vytvoriť nový Jupyter Notebook vo VS Code. Prvým je vytvorenie nového súboru, ktorý je pomenovaný s koncovkou .ipynb, v adresári, ktorý je nastavený ako pracovný priestor. Druhým spôsobom je spustenie príkazu „Create: New Jupyter Notebook“ z palety príkazov v hornej lište VS Code.

Nasleduje návod so snímkami obrazovky, ako vybrať a nastaviť kernel. VS Code ponúka 3 možnosti: Python prostredia, Jupyter kernel a existujúci Jupyter server. Pre túto prácu je podstatná prvá možnosť.

Markdown bunky

V štvrtej časti sú popísané a vysvetlené markdown bunky. Nasledujú ukážky a jednoduché cvičenia na vytváranie nadpisov, zoznamov, zvýrazňovanie textu, vytváranie hyperlinkov na URL (Uniform Resource Locator) adresy a aj v rámci notebooku (napr. odkaz na nadpis). Vďaka týmto cvičeniam si užívateľ vyskúša a osvojí základné zručnosti práce s markdown bunkami. Nadpisy vedú vytvárať hierarchickú štruktúru buniek, ktoré umožnia definovať jednotlivé kapitoly a podkapitoly notebooku. Tieto sa dajú v prípade potreby zbalíť. Na záver sú vysvetlené 2 spôsoby, ako vkladať obrázky do markdown buniek. Prvým je kombinácia kopírovať + prilepiť, vtedy je obrázok vložený ako príloha (attachement). Druhým je použitie HTML tagu , kde sa píše cesta k obrázku

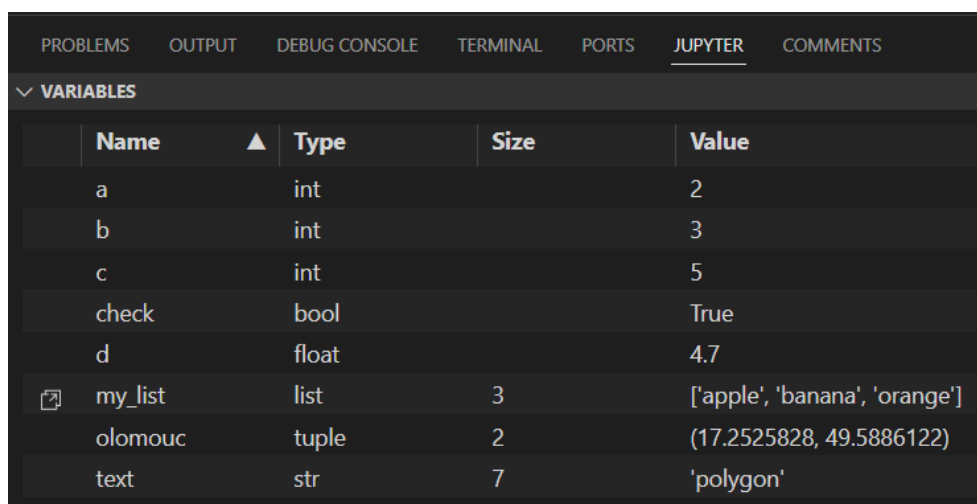
uloženému v sieťovom prostredí (internet) alebo v lokálnom úložisku počítača. Tento spôsob umožňuje meniť výšku a šírku obrázka.

Bunky kódu

Táto časť popisuje bunky kódu a uvádza príklady jednoduchých Python príkazov na výpis textu, sčítanie, odčítanie, vytváranie premenných. Premenná vytvorená v jednej bunke kódu môže byť použitá v akejkolvek inej bunke v notebooku. Premenné sa ukladajú do pamäte a sú v nej dostupné až do zatvorenia notebooku. V notebooku je možné vypísať hodnotu premennej 2 spôsobmi, prvým je príkaz `print` a druhým je napísanie názvu premennej. Príkaz `print` je v určitom zmysle „istejší“, pretože vždy vypíše to, čo má. Napísanie názvu premennej vypíše jej hodnotu iba vtedy, ak sa za týmto riadkom kódu nenachádza ďalší riadok s príkazom na výpis.

Ďalšou témou, ktorej je venovaných pár buniek kódu, je dynamické typovanie. V Pythone je možné do premennej nahráť hodnotu jedného dátového typu a v ďalšom riadku je možné do tej istej premennej uložiť hodnotu iného dátového typu. Dátový typ premennej teda nie je fixný a v priebehu skriptu/notebooku sa môže meniť. VS Code ponúka prehľad vytvorených premenných (Obr. 12), čo je tabuľka s názvami, typmi, veľkosťou a hodnotami premenných.

Posledné ukážky v tejto časti sa venujú rozdeľovaniu a spájaniu buniek v notebooku.



Name	Type	Size	Value
a	int		2
b	int		3
c	int		5
check	bool		True
d	float		4.7
my_list	list	3	['apple', 'banana', 'orange']
olomouc	tuple	2	(17.2525828, 49.5886122)
text	str	7	'polygon'

Obr. 12 Prehľad vytvorených premenných vo VS Code (autor: A. Tóth).

Užitočné Python príkazy

Na záver notebooku je vypísaných a vysvetlených niekoľko Python príkazov a funkcií, ktoré je potrebné poznať na úspešné zvládnutie cvičení v nasledujúcom notebooku Python cvičenia. Odkazy v názvoch metód, príkazov a funkcií smerujú na W3Schools Python tutoriál (W3Schools 2024):

- [input](#) – po spustení si pýta textový vstup od užívateľa.
- [modulus](#) – vráti zvyšok po celočíselnom delení.
- [F-refazec](#) – používa sa pri výpise, keď je potrebné kombinovať text s hodnotami premenných. Text sa napíše do úvodzoviek a premenné sa v rámci tohto textu vložia do zložených zátvoriek.

- [range](#) – definuje číselný rozsah, používa sa napr. v hlavičke for cyklov, je možné definovať začiatok rozsahu, koniec rozsahu a krok, o ktorý sa má napr. cyklus v rozsahu posúvať.
- [split](#) – rozdelí textový reťazec (string) na miestach, kde sa nachádza zadaný znak a rozdelené časti reťazca uloží do zoznamu. Typickým príkladom je veta, ktorá je funkciou split rozdelená na slová (zadaný znak je medzera) a tieto slová sú prvkami zoznamu.
- [randint](#) – funkcia vstavaného Python modulu random, vygeneruje náhodné celé čísla z rozsahu definovaného v zátvorkách funkcie. Do rozsahu spadajú obe hraničné hodnoty.

4.1.2 Python cvičenia

Názov súboru: **1_python_exercises.ipynb**

Potrebný komerčný software: nie

Počet markdown buniek: 19

Počet buniek kódu: 8

Počet riadkov kódu: 70

Tento notebook obsahuje 8 cvičení, ktoré precvičia základné Python zručnosti a programátorské myslenie. Notebook je rozdelený na dve časti: v prvej sú zadania a v druhej riešenia. Existuje viac správnych riešení, tie ktoré sú uvedené, sú len jednou z možností. Zadania cvičení boli zostavené s pomocou technológie ChatGPT 3.5 (OpenAI 2023), vygenerovaná odpoveď je na Obr. 13. Riešenia boli vypracované samostatne autorom tejto práce.



You

Give me some basic Python exercises for students.



ChatGPT

Sure, here are some basic Python exercises for students:

1. Write a program that asks the user to enter their name and then prints "Hello, [name]!".
2. Write a program that takes a number as input from the user and prints whether it is even or odd.
3. Write a program that takes two numbers as input from the user and prints their sum.
4. Write a program that takes a string as input from the user and prints it in reverse.
5. Write a program that generates a random number between 1 and 10 and asks the user to guess the number. The program should give feedback on whether the guess was too high or too low, and keep asking for guesses until the correct number is guessed.
6. Write a program that takes a list of numbers as input from the user and prints the sum of the numbers.
7. Write a program that takes a sentence as input from the user and counts the number of words in the sentence.
8. Write a program that takes a list of numbers as input from the user and prints the largest and smallest numbers in the list.
9. Write a program that takes a string as input from the user and prints the number of vowels in the string.
10. Write a program that takes two lists of numbers as input from the user and prints the elements that are common to both lists.

Obr. 13 Zadania Python cvičení vygenerované technológiou ChatGPT 3.5 (autor: A. Tóth).

- zadanie:* Napíšte kód, ktorý si od užívateľa funkciou input vypýta celé číslo a potom vypíše, či je zadané číslo párne alebo nepárne.

riešenie: Do premennej number sa uloží číslo z funkcie input a nasleduje podmienka, ktorá vypíše vetu „číslo je párne“ vtedy, keď sa zvyšok po celočíselnom delení zadaného čísla dvomi sa rovná 0. Veta „číslo je nepárne“ je vypísaná v opačnom prípade (vetva else).
- zadanie:* Upravte predchádzajúci kód tak, že vstupom bude zoznam 6 celých čísel namiesto jedného čísla zadaného užívateľom a program vypíše pre každé číslo zo zoznamu, či je párne alebo nepárne.

riešenie: Na začiatku je vytvorený zoznam so 6 ľubovoľnými celými číslami a nasleduje for cyklus, ktorý prechádza zoznam a každé číslo vstupuje do podmienky definovanej v predchádzajúcom cvičení.
- zadanie:* Napíšte kód, ktorý si funkciou input od užívateľa vypýta jeho meno a vypíše ho odzadu

riešenie: Meno z inputu je uložené do premennej name. Do premennej name_backward sa uloží prázdny reťazec. Vo for cykle, ktorý iteruje v rozsahu range(len(name)-1, -1, -1), sa do premennej name_backward ukladá znak po znak meno odzadu. Prvým písmenom, ktoré sa uloží do tejto premennej, je posledné písmeno mena, pretože cyklus začína na pozícii len(name)-1, čo je pozícia posledného znaku. Cyklus končí na pozícii 0, čo je pozícia prvého písmena, takže posledné uložené písmeno v premennej name_backward je prvé písmeno mena. V rozsahu je tiež definovaný krok -1, čo znamená, že postupuje dozadu, od posledného písmena k prvému. Keď cyklus skončí, je vypísaná premenná name_backward s menom odzadu.
- zadanie:* Vymyslite si vetu, uložte ju do premennej sentence, zistite počet slov tejto vety a vypíšte ho.

riešenie: Do premennej sentence je uložená ľubovoľná veta, ktorá je funkciou split podľa deliaceho znaku medzera rozdelená na slová. Tieto slová sú uložené v zozname words_list. Na konci je vypísaný počet slov ako dĺžka zoznamu words_list.
- zadanie:* Spočítajte počet samohlások v názve hlavného mesta Uruguaja a vypíšte tento počet.

riešenie: Do premennej word je uložený reťazec „Montevideo“ a do premennej vowels je uložený reťazec samohlások „aeiouy“. Premenná count je nastavená na hodnotu 0. For cyklom je prechádzaný reťazec „Montevideo“ znak po znaku a ak sa aktuálny znak nachádza v reťazci samohlások, hodnota v premennej count sa zvýši o 1. Po skončení cyklu je počet samohlások v reťazci „Montevideo“ uložený v premennej count, ktorá je vypísaná.
- zadanie:* Vytvorte si 2 zoznamy čísel, porovnajte ich, uložte spoločné čísla do nového zoznamu a tento zoznam vypíšte.

riešenie: Vytvorené sú 2 zoznamy, list_a a list_b, každý obsahuje 6 čísel a niektoré sa nachádzajú v oboch zoznamoch. Tretí zoznam common je zatiaľ prázdny. Prvý

for cyklus premennou a iteruje cez list_a. V tomto for cykle je vnorený druhý for cyklus, ktorý premennou b iteruje cez list_b. Ak sa a rovná b, pridaj b do zoznamu common, do ktorého sa ukladajú spoločné čísla. Takto sa každé číslo zo zoznamu list_a porovná s každým číslom zo zoznamu list_b a všetky spoločné čísla sa uložia do zoznamu common, ktorý je skončení oboch cyklov vypísaný.

7. *zadanie:* Napíšte program, ktorý vygeneruje náhodné číslo z intervalu 1-10 a vyzve užívateľa, aby hádal, aké číslo bolo vygenerované. Program dá spätnú väzbu, či bol tip správny alebo nie a ak nebol správny, napíše, či je hádané číslo väčšie alebo menšie ako to vygenerované. Program beží a pýta sa užívateľa dovedy, dokým neuhádne vygenerované číslo.

riešenie: Modul random je importovaný príkazom import random. Do premennej r sa uloží funkciou randint náhodne vygenerované číslo z intervalu 1-10. Do premennej g sa budú ukladať užívateľove tipované čísla. Pomocou funkcie input je užívateľ vyzvaný k prvému tipu. Ak uhádol (splnená podmienka že g sa rovná r), je vypísaná hláška, že uhádol na prvýkrát a program končí. Ak neuhádol, spúšťa sa while cyklus, ktorý beží dovedy, kým sa g nerovná r. V tomto cykle sa vypisuje spätná väzba, či bolo hádané číslo väčšie alebo menšie ako to náhodne vygenerované. Iterácia cyklu je zakončená výzvou na ďalší tip. Keď užívateľ uhádne číslo, cyklus sa skončí a vypíše sa hláška o úspešnom uhádnutí a celý program končí.

8. *zadanie:* Vypíšte prvých 12 čísel Fibonacciho postupnosti, pričom prvé 2 čísla sú 0 a 1.

riešenie: Do premennej count sa uloží číslo 12 a do zoznamu fib sa uložia prvé 2 čísla, ktorými sú 0 a 1. Následne sa spustí for cyklus, ktorý premennou i iteruje od 2 do 12 a do zoznamu fib pridáva číslo, ktoré je súčtom predchádzajúcich 2 prvkov zoznamu fib ($fib[i-1] + fib[i-2]$). Takto sa do zoznamu fib uloží prvých 12 čísel Fibonacciho postupnosti, ktoré sú po skončení for cyklu vypísané. Sú to tieto čísla: 0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89.

4.1.3 ArcPy cvičenia

Názov súboru: **2_arcpy_exercises.ipynb**

Potrebný komerčný software: áno, ArcGIS Pro

Počet markdown buniek: 38

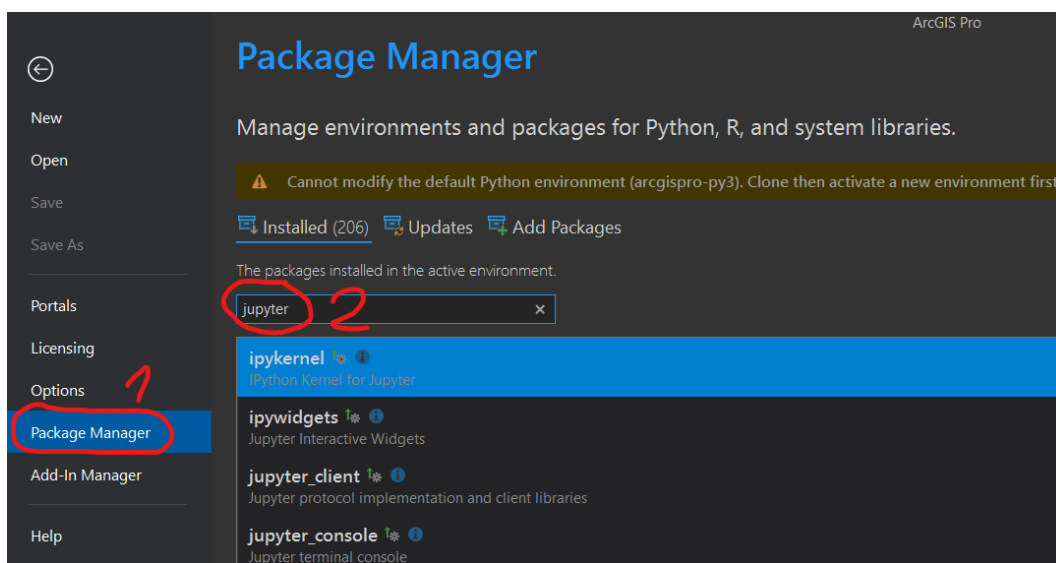
Počet buniek kódu: 11

Počet riadkov kódu: 59

Tento notebook obsahuje základné cvičenia na funkcie knižnice ArcPy, ktorými si užívateľ osvojí prácu s touto knižnicou v Jupyter Notebooku. Tento notebook je rozdelený na 3 časti: prvá obsahuje návod na naklonovanie ArcGIS Python prostredia, v druhej je ukázané, ako nastaviť toto prostredie ako kernel vo VS Code a tým pádom používať ArcGIS funkcionálnosť mimo ArcGISu a tretia časť je sprievodcom, ako používať Jupyter Notebooky v ArcGIS Pro, čím sa stávajú ArcGIS Notebookmi. Prerekvizitami na úspešné používanie tohto notebooku je licencia na ArcGIS Pro, funkčný ArcGIS Pro nainštalovaný v počítači a dátová sada ArcCR 500 verzia 3.3.

ArcGIS Python prostredie

Python je nedeliteľnou súčasťou ArcGISu, na Pythone sú postavené napr. nástroje geoprocessingu. Pri inštalácii ArcGIS Pro je stiahnutý a inštalovaný aj Python s knižnicami a balíkami, ktoré sú potrebné pre správne fungovanie ArcGIS Pro. Tento Python vytvára základné prostredie, tzv. „default Python environment“. Náhľadom do tohto prostredia je záložka Package manager v nastaveniach ArcGIS Pro. Tam je zoznam nainštalovaných knižníc a po napísaní „jupyter“ do vyhľadávacieho riadku sa vyfiltrujú všetky knižnice súvisiace s Jupyter Notebookmi. Prvá v zozname je ipykernel, čo je knižnica umožňujúca spúšťanie kódu v notebooku. Takže ArcGIS Python prostredie natívne podporuje používanie Jupyter Notebookov. V žltom obdĺžniku nad zoznamom knižníc je zvýraznené varovanie, že toto základné prostredie nie je možné upravovať. Na inštaláciu ďalších knižníc je potrebné naklonovať toto základné prostredie a aktivovať ho (Obr. 14).



Obr. 14 Ukážka ArcGIS Python prostredia a nainštalovaných knižníc podporujúcich prácu s Jupyter Notebookmi (autor: A. Tóth).

V prípade, že užívateľ nepotrebuje inštalovať dodatočné knižnice a s notebookmi chce pracovať výlučne v ArcGIS Pro, prostredie si nemusí klonovať. V opačnom prípade je klonovanie nevyhnutné. Nasleduje návod, vytvorený pomocou snímok obrazovky, ako si naklonovať a aktivovať toto prostredie.

Inštalovať knižnice do naklonovaného ArcGIS Python prostredia sa dá viacerými spôsobmi. Jedným z nich je už spomenutý Package manager v nastaveniach ArcGIS Pro. Naklonované prostredie musí byť nastavené ako aktívne. V záložke Add Packages je vyhľadaná želaná knižnica a ak sa nachádza v zozname, po kliknutí na ňu sa v paneli napravo zobrazia informácie o nej spolu s tlačidlom inštalovať. Druhým spôsobom je otvoriť Python príkazový riadok (Python command prompt), ktorým je možné prísť k aktívnemu prostrediu. Príkazom `pip install + názov knižnice` sa nainštaluje želaná knižnica.

Jupyter Notebooky vo VS Code s ArcGIS Python prostredím

Po úspešnom dokončení klonovania a aktivácie prostredia si užívateľ založí nový projekt ArcGIS Pro. Potom nastaví naklonované Python prostredie ako kernel vo VS Code a môže začať používať funkcionality ArcPy (Esri no date) v nasledujúcich cvičeniach.

1. *zadanie:* Importujte knižnicu ArcPy, nastavte geodatabázu vášho ArcGIS projektu ako pracovný priestor a povoľte prepisovanie výstupov s rovnakým názvom.
riešenie: Príkaz import arcpy je na prvom riadku, nasleduje uloženie cesty ku geodatabáze do [arcpy.env.workspace](#) a [arcpy.env.overwriteOutput](#) je nastavený na hodnotu True.
2. *zadanie:* Uložte si cestu ku geodatabáze AdministrativniCleneni_v13.gdb, exportujte obec Lutín z vrstvy ObceBody zo spomínanej geodatabázy do samostatnej vrstvy vo vašej geodatabáze. Pomocou metódy ListFeatureClasses zistite, či export prebehol.
riešenie: Do premennej arc_admin je uložená cesta ku geodatabáze AdministrativniCleneni_v13.gdb a pomocou funkcie [ExportFeatures](#) je obec Lutín exportovaná do novej vrstvy lutin_point. Po zavolaní metódy [ListFeatureClasses](#) sú vypísané všetky vrstvy v geodatabáze, ktorá bola nastavená ako pracovný priestor. V tomto okamihu by tam mala byť iba jedna vrstva lutin_point.
3. *zadanie:* Vytvorte 10 km široký buffer (obalová zóna) okolo exportovanej obce Lutín.
riešenie: Buffer je vytvorený ArcPy funkciou [Buffer](#), do prvého parametra je vložená vrstva obce Lutín, druhý parameter je názov výstupnej vrstvy a tretím parametrom je veľkosť obalovej zóny.
4. *zadanie:* Z geodatabázy ArcCR500_v33.gdb orežte vrstvy Lesy, SidlaPlochy, Silnice_2016, VodniToky vytvoreným 10 km bufferom. Z geodatabázy AdministrativniCleneni_v13.gdb orežte vrstvu Obce – Polygony tým istým bufferom.
riešenie: Do premennej arc je uložená cesta ku geodatabáze ArcČR 500 a do zoznamu arc_layers sú uložené názvy vrstiev, ktoré majú byť orezané. For cyklus iteruje cez tento zoznam a každá vrstva je orezaná 10 km bufferom pomocou funkcie [Clip](#). Ak má vrstva názov „ObcePolygony“, do prvého parametra funkcie Clip je vložená cesta ku geodatabáze s administratívnym členením, kde sa táto vrstva nachádza. Všetky orezané vrstvy majú na konci názvu príponu „_cl“

Jupyter Notebooky v ArcGIS Pro

Po prejdení 4 cvičení je čas opustiť VS Code a vyskúšať pracovať s notebookom v ArcGIS Pro. Tento notebook je pridaný do vytvoreného projektu. Návod ako pridať existujúci Jupyter Notebook do projektu ArcGIS Pro a otvoriť ho je zdokumentovaný pomocou snímok obrazovky. Rozhranie notebooku je iné ako rozhranie vo VS Code, ale princíp zostáva rovnaký. Užívateľ môže pokračovať v ďalších 4 cvičeniach, najprv však ale musí spustiť kód z prvého zadania, aby sa importovala knižnica ArcPy a nastavil sa pracovný priestor.

5. *zadanie:* Vypíšte geometrický typ a počet prvkov vo všetkých 5 orezaných vrstvách. Môžete použiť nástroj Get Count a vlastnosti vrstvy z funkcie Describe.
riešenie: Riešením je jeden for cyklus, ktorý iteruje cez všetky vrstvy v pracovnom priestore (geodatabáza projektu), ktoré sa končia znakmi „_cl“. Tento výber orezaných vrstiev je uskutočnený funkciou List Feature Classes. Prvým riadkom vo for cykle je zistenie počtu prvkov vrstvy pomocou funkcie [Get Count](#). Funkcia [Describe](#) vytvorí objekt, ktorý obsahuje vlastnosti vrstvy. Nasledujú 4 príkazy print, ktoré postupne vypíšu názov aktuálnej vrstvy, počet jej prvkov, geometrický typ

pomocou vlastnosti Describe objektu [shapeType](#) a na záver prázdny reťazec, aby sa výpis ďalšej vrstvy viditeľne oddelil.

6. *zadanie:* Získajte celkovú dĺžku všetkých ciest z orezanej vrstvy Silnice_2016_cl. Vypíšte túto dĺžku v km a zaokrúhlenú na 2 desatinné miesta.

riešenie: Premenná length bude obsahovať celkovú dĺžku ciest, na začiatku je nastavená na hodnotu 0. Metóda [Search Cursor](#) vytvorí iterovateľný objekt cursor s 2 atribútmi OBJECTID a SHAPE_Length vrstvy Silnice_2016_cl. Cez tento cursor sa vo for cykle prístupuje k dĺžkam jednotlivých úsekov ciest a premenná length sa tak každou iteráciou zväčší o dĺžku aktuálnej cesty. Po skončení cyklu je vypísaná dĺžka, ktorá je vydelená 1000 (aby bola v km) a zaokrúhlená funkciou round na dve desatinné miesta.

7. *zadanie:* Nájdite najdlhšiu cestu v tej istej vrstve Silnice_2016_cl a vypíšte jej dĺžku a OBJECTID.

riešenie: Premenná maxlength bude obsahovať dĺžku najdlhšej cesty a v maxid bude uložené OBJECTID tejto cesty. Obe sú na začiatku nastavené na 0. Nasleduje metóda Search Cursor s rovnakým nastavením ako pri predchádzajúcom cvičení a for cyklus znovu prechádza cez jednotlivé cesty vrstvy Silnice_2016_cl. V tomto cykle je podmienka, ktorá kontroluje, či je dĺžka aktuálnej cesty väčšia ako hodnota v maxlength. Ak áno, tak premennej maxlength je priradená táto dĺžka a premennej maxid je priradené OBJECTID tejto cesty. Po skončení cyklu sú vypísané obidve premenné.

8. *zadanie:* Vypočítajte hustotu ciest ako počet km ciest na 1 km² plochy pre každú obec z vrstvy ObcePolygony_cl. Zapište túto hustotu do nového atribútu s názvom rd_density. Nároveň – použite nástroje [Intersect](#), [Dissolve](#), [Join Field](#), [Add Field](#) a [Calculate Field](#).

riešenie: Nástrojom Intersect je vrstva ciest rozsekaná podľa polygónovej vrstvy obcí. Nástrojom Dissolve je rozsekaná vrstva upravená tak, že všetky cesty v jednej obci sú spojené do jedného prvku. Geometria ciest sa nemení, mení sa iba počet záznamov v atribútovej tabuľke. Následne je v tejto vrstve vytvorený nový atribút rd_length, do ktorého je uložená dĺžka ciest (v podstate nakopírovaný atribút SHAPE_Length) pomocou funkcie Calculate Field. Atribút rd_length je funkciou Join Field pripojený k vrstve obcí. V tejto vrstve je potom do nového atribútu rd_density vypočítaná hustota ciest ako ich dĺžka v km vydelená rozlohou v km².

4.2 Kolokačná analýza

Umiestnenie a názov súboru: /1_colocation_analysis/colocation_analysis.ipynb

Vstupné dáta: Urban Atlas Land Use/Land Cover 2018

Konkrétny prípad: výskyt zelene a športových zariadení v blízkosti mestskej zástavby v konurbácii Katowice, Poľsko

Potrebný komerčný software: áno, ArcGIS Pro

Počet markdown buniek: 79

Počet buniek kódu: 50

Počet riadkov kódu: 270

Úvod

Notebook sa venuje kolokačnej analýze, vysvetľuje, čo to je, ako funguje. Je sprievodcom pre tých, ktorí chcú pochopiť, ako kolokačná analýza funguje, čo vstupuje do výpočtu kolokačného koeficientu, ako jednotlivé parametre ovplyvňujú výsledok analýzy. Ponúka možnosť spustiť si nástroj [Colocation analysis](#) s rôznymi parametrami a graficky porovnať výsledky. Je rozdelený na 6 častí, pričom tieto na seba nadväzujú.

Notebook je šírený spolu s projektom ArcGIS Pro, s ktorým má byť používaný. Projekt je zabalený vo formáte project package (colocation_agp_project.ppkx) a sú v ňom vytvorené bookmarky, ktoré slúžia ako pomocník pri priestorovej interpretácii výsledkov kolokačnej analýzy a je na nich postavená jedna celá časť notebooku. Súčasťou projektu je aj geodatabáza s dvoma vrstvami. Prvou je polygónová vrstva UA LULC (Urban Atlas Land Use/Land Cover), ktorá je zároveň pridaná do mapy a má nastavenú originálnu Urban Atlas symbológiu. Druhá vrstva obsahuje polygón Urban Core, ktorý ohraničuje skúmanú oblasť. Ak by si užívateľ samostatne stiahol dáta z Urban Atlasu a vytvoril by si vlastné bookmarky, mohol by používať a spúšťať tento notebook v akomkoľvek inom projekte ArcGIS Pro. Na stiahnutie dát z Urban Atlasu je potrebná bezplatná registrácia v [Copernicus Land Monitoring Service](#).

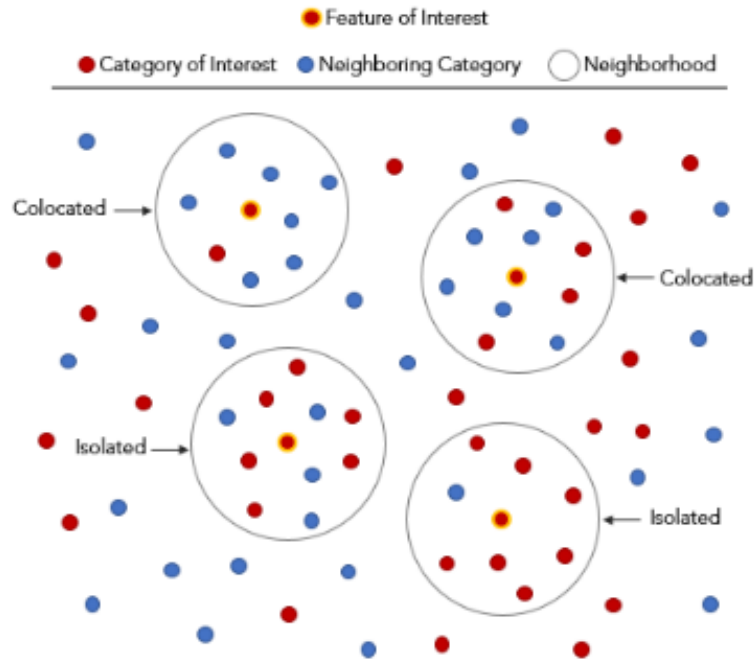
Kernelom tohto notebooku je naklonované ArcGIS Python prostredie, do ktorého je potrebné nainštalovať knižnice Pandas (pandas via NumFOCUS Inc. 2024) a Matplotlib (The Matplotlib development team 2024).

Užívateľ môže slobodne prepínať medzi panelom mapy a notebooku alebo si môže pracovnú ArcGIS plochu rozdeliť na polovicu medzi tieto 2 panely. Prepínanie medzi panelmi má výhodu v tom, že je vidieť vždy celá mapa a celý notebook, ale vyžaduje akciu navyše, ktorou je pohnutie myšou a kliknutie. Rozdelenie plochy na polovicu síce obmedzuje šírku notebooku a mapy, takže ak sú riadky kódu dlhšie, nie sú vidno naraz celé, ale na druhej strane užívateľ hneď vidí výsledky ArcPy funkcií v mape a nemusí pre to spraviť ďalšiu akciu. Prepojenie Jupyter Notebooku a ArcGISu teda umožňuje tzv. „interaktívne“ spúšťanie Python kódu.

Čo je kolokačná analýza?

Kolokačná analýza „meria lokálne vzory/vzorcie priestorovej asociácie – kolokácie – medzi 2 kategóriami bodových prvkov pomocou štatistiky kolokačného koeficientu“ (Esri no date). Inými slovami, kolokačná analýza zisťuje, či má jedna kategória tendenciu vyskytovať sa v blízkosti druhej kategórie – kolokovať – a aká veľká je táto tendencia. Vstupom do tejto analýzy je bodová vrstva, ktorej prvky sú rozdelené do 2 kategórií: kategória záujmu (body A) a susediaca kategória (body B). Nasleduje skúmanie okolia každého bodu A a určenie, či je tento bod A kolokovaný s bodmi B alebo naopak izolovaný od bodov B (Obr. 15).

Miera kolokácie alebo izolácie sa meria pomocou LCLQ (lokálny kolokačný koeficient). Tento koeficient má pomerne komplexný vzorec, podľa ktorého je počítaný, vo všeobecnosti ide o porovnanie, resp. delenie podielu bodov B v okolí daného bodu A podielom bodov B v celej skúmanej oblasti. Ak je lokálny podiel bodov B väčší ako globálny podiel, LCLQ bude väčší ako 1 a daný bod A bude označený ako kolokovaný. Ak je lokálny podiel bodov B menší ako globálny podiel, LCLQ bude tým pádom menšie ako 1 a daný bod A bude izolovaný. Lokálny podiel bodov B môže byť ovplyvňovaný váhami, kedy najbližšiemu susediacemu bodu je priradená najvyššia váha a najvzdialenejšiemu bodu zase najmenšia.

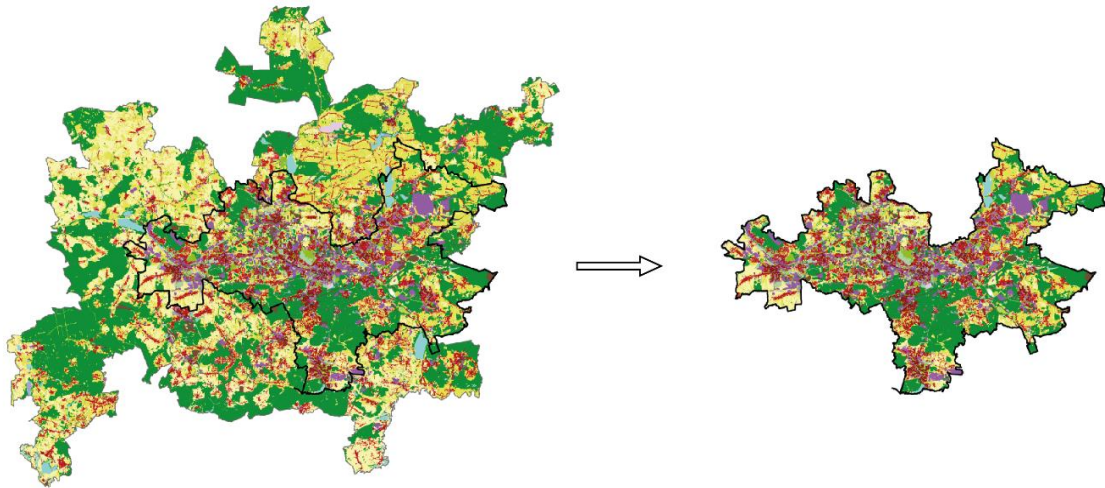


Obr. 15 Jednoduchá ilustrácia výsledku kolokačnej analýzy (zdroj: [How Colocation Analysis works](#)).

Kolokačná analýza je v prostredí ArcGIS Pro realizovaná nástrojom Colocation Analysis z toolboxu Spatial Statistic. Tento nástroj umožňuje prispôbiť analýzu pre konkrétne prípady, poskytuje viaceré nastavenia, ktoré si užívateľ môže zvoliť, ako napr. rôzne definície okolia (fixná vzdialenosť verzus k najbližším susedov), váhové schémy priraďujúce váhy bodom v okolí alebo počet permutácií, ktoré počítajú p-value. Pre bežného užívateľa však môže byť množstvo a variabilita vstupných parametrov náročná na pochopenie alebo môže byť ťažko predstaviteľné, aký vplyv má daný parameter na výsledok analýzy. Tento notebook pomôže pochopiť jednotlivé parametre a v istom zmysle „pozrieť sa na kolokačnú analýzu zblízka“ na konkrétnom príklade v konurbácii Katowice v Poľsku. Kategóriou záujmu budú body reprezentujúce mestskú zástavbu a susediacou kategóriou budú body reprezentujúce zeleň a športoviská. Predmetom skúmania je odhaliť miesta, kde majú obyvatelia v blízkosti bydliska alebo práce viac priestoru na relax vo forme pohybu alebo pobytu v zelenom prostredí a miesta, kde je takýchto možností menej.

Príprava dát na analýzu a nastavenie pracovného priestoru

V tejto prvej časti notebooku sú najprv nastavené premenné ArcPy prostredia ako `arcpy.env.workspace` a `arcpy.env.overwriteOutput`, uloží sa prístup k aktuálnemu ArcGIS projektu a mape, čo je v ňom, aby bolo možné s touto mapou a vrstvami v nej manipulovať. Ďalej je vrstvou Urban Core orezaná vrstva UA LULC, nakoľko je táto vrstva stiahnutá pre celú katowickú FUA, ktorá zahŕňa aj vidiecke prostredie v blízkosti katowickej konurbácie. V tomto prípade je oblasťou záujmu výlučne mestské prostredie z dôvodu skúmanej priestorovej asociácie – kolokácie, a preto musí byť vstupná vrstva orezaná tak, aby obsahovala mestá bez vedľajších vidieckych sídel (Obr. 16).



Obr. 16 Ilustrácia orezania vrstvy UA LULC na mestské jadro – Urban Core (autor: A. Tóth).

Následne je orezaná vrstva pridaná do mapy funkciou [Add Data From Path](#) a pomocou [Apply Symbology From Layer](#) je na ňu aplikovaná originálna Urban Atlas symbolológia, ktorá je uložená v pôvodnej neorezanej vrstve.

Nasleduje agregácia kategórií v UA LULC do skupín. Skúmaná kolokačná otázka znie: vyskytujú sa zeleň a športoviská v blízkosti mestskej zástavby? Z tejto otázky vyplývajú ďalšie: čo je zeleň, športoviská a mestská zástavba? Ako sú definované v UA LULC? Aká vzdialenosť je považovaná za blízkosť?

UA LULC je polygónová vrstva obsahujúca 25 kategórií využitia územia (land use). Kategórie na Obr. 17 vhodne reprezentujú zeleň a športoviská a preto sú zlúčené do jednej skupiny s pracovným názvom „relax“.

- 14100: Green urban areas
- 14200: Sports and leisure facilities

Obr. 17 Kategórie zelene a športovísk vo vrstve UA LULC (autor: A. Tóth, zdroj: Urban Atlas 2018 delivery report).

Kategórie, ktorých kód začína číslom 11***, sú kategórie mestskej zástavby () a preto je z nich vytvorená skupina „urban fabric“.

- 11100: Continuous urban fabric (S.L.: > 80%)
- 11210: Discontinuous dense urban fabric (S.L.: 50% - 80%)
- 11220: Discontinuous medium density urban fabric (S.L.: 30% - 50%)
- 11230: Discontinuous low density urban fabric (S.L.: 10% - 30%)
- 11240: Discontinuous very low density urban fabric (S.L.: < 10%)
- 11300: Isolated structures

Obr. 18 Kategórie mestskej zástavby vo vrstve UA LULC (autor: A. Tóth, zdroj: Urban Atlas 2018 delivery report).

Ostatné kategórie sú zoskupené tiež podľa prvých 2 alebo 3 čísel v ich kódach, niektoré kategórie tvoria skupiny samostatne. Pre potreby tohto notebooku zoskupenie ostatných kategórií nebolo nakoniec potrebné, pretože pre prehľadnosť a primeraný rozsah notebooku bolo upustené od výberu dvojice skupín pre kolokačnú analýzu na základe CLQ (globálneho kolokačného koeficientu). Tento koeficient kvantifikuje kolokáciu všetkých dvojíc skupín vo vstupnej vrstve na globálnej úrovni, čiže v celom skúmanom území. Dvojica skupín s vysokým alebo nízkym CLQ by následne bola vstupom pre kolokačnú analýzu, ktorá hľadá lokálne vzory kolokácie.

Nástroj Colocation analysis potrebuje na vstupe bodovú vrstvu, takže z pôvodnej vrstvy sú vybrané polygóny skupín „urban_fabric“ a „relax“ a z nich sú vygenerované centroidy, ktoré budú vstupovať do analýzy. Pri tomto postupe boli použité funkcie [Select Layer By Attribute](#) a [Feature To Point](#). Po dohode s vedúcou práce sú centroidy generované tak, aby boli vnútri príslušného polygónu, hoci to môže znamenať posunutie z presného geometrického stredu polygónu.

Nastavenia kolokačnej analýzy

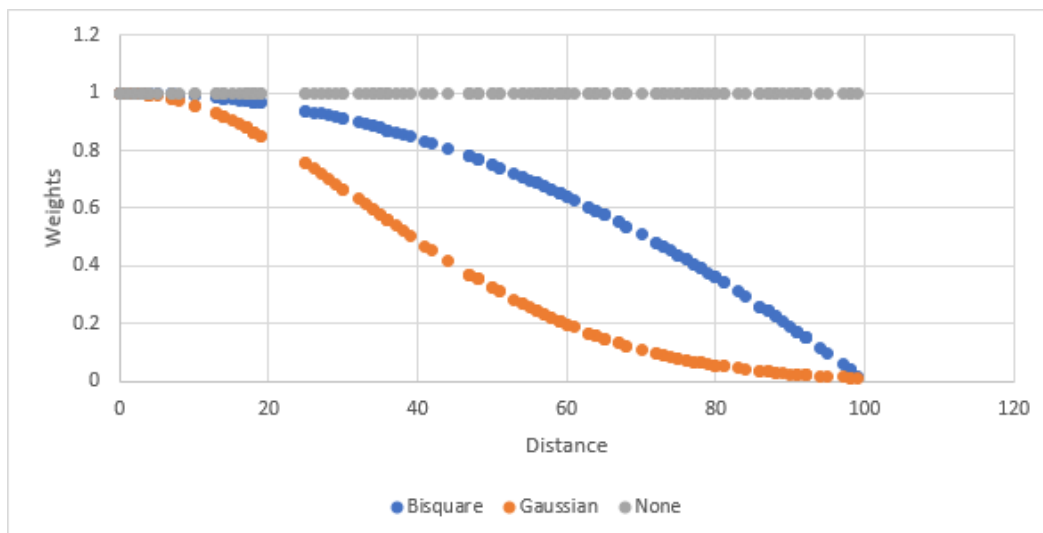
Druhou časťou notebooku je vysvetlenie a nastavenie jednotlivých parametrov nástroja Colocation analysis a samotné spustenie analýzy.

Parametru input_type je priradená hodnota single dataset, čo znamená, že obidve skupiny sú v jednom datasete, resp. v jednej vrstve. V ďalších parametroch je špecifikovaná vrstva centroidov ako vstupná vrstva, „urban_fabric“ ako kategória záujmu a „relax“ ako susediaca kategória. Do parametru field_of_interest je vložené atribútové pole obsahujúce informáciu a tom, či je daný centroid reprezentuje skupinu „urban_fabric“ alebo „relax“.

Okolie je v tomto prípade definované fixnou vzdialenosťou (distance band), ktorá bola po dohode s vedúcou práce stanovená na 800 m. "Hodnota 800 m je priemerná vzdialenosť troch blokov v UA LULC datasete. Bola zvolená ako vhodná maximálna vzdialenosť pre dobrú zmes meniacej sa hustoty blokov“. Všetky body (centroidy) nachádzajúce sa do vzdialenosti 800 m od skúmaného bodu záujmu budú považované za susedov. Susedia potom vstupujú do výpočtu LCLQ. Okolie môže byť definované tiež počtom susedov, čiže za susedov bude považovaných napríklad 8 najbližších bodov ku skúmanému bodu záujmu.

Dôležitým nastavením je určenie počtu permutácií, pomocou ktorých sa počíta p-value pre každý bod kategórie záujmu. P-value určuje, či je LCLQ pre daný bod štatisticky významný alebo nie. Pri každej permutácii sa kategórie bodov v celom území „zamiešajú a náhodne rozdeľujú“ – odborným výrazom permutujú. Poloha bodov ostáva stála, mení sa iba kategória, takže bod pri jednej permutácii môže byť kategória A, pri druhej B, pri tretej znovu B, pri štvrtej A atď. Pri každej permutácii sa vypočíta LCLQ pre každý bod kategórie A. Výsledkom permutácií je referenčné rozdelenie hodnôt LCLQ, ktoré sa následne porovná s vypočítanou hodnotou LCLQ daného bodu, aby sa určila pravdepodobnosť, s akou je táto hodnota výsledkom náhodného procesu. Táto pravdepodobnosť sa nazýva p-value a ak je menšia ako 0.05, tak LCLQ daného bodu nie je výsledkom náhody a je štatisticky významný (Esri no date). Nástroj umožňuje vybrať z 5 počtov permutácií: 99, 199, 499, 999, 9999. Výber počtu permutácií je balansovaním medzi presnosťou p-value a nárokmi na výkon počítača a výpočtový čas (Esri no date). Čím viac permutácií, tým robustnejšie a presnejšie referenčné rozdelenie, ktoré určuje hodnotu p-value. V tomto prípade bola zvolená prostredná možnosť, čiže 499 permutácií.

Lokálna váhová schéma definuje váhu, akou vplývajú susedné body na aktuálny/daný bod pri výpočte LCLQ. Na výber sú 3 možnosti: Bisquare, Gaussian a None. Bisquare a Gaussian priradujú susedným bodom váhy na základe ich vzdialenosti k aktuálnemu bodu, rozdiel je v tom, že váhy v Gaussian schéme klesajú rýchlejšie s narastajúcou vzdialenosťou ako pri Bisquare schéme, ako ilustruje Obr. 19.



Obr. 19 Lokálne váhové schémy v nástroji Colocation analysis (zdroj: [How Colocation Analysis works](#)).






Možnosť None nenastavuje žiadne váhy, resp. všetky susedné body majú rovnakú váhu a nezáleží na vzdialenosti k aktuálnemu bodu. Viac o lokálnych váhových schémach je napísané v článku „How Colocation Analysis works“ (Esri no date). Gaussian váhová schéma bola vybraná pre tento prípad.

Posledným nastavením je pomenovanie výstupnej vrstvy. Názov „ca_ur_ab_800_g_499“ obsahuje viacero skratiek a hodnôt vstupných parametrov. „Ca“ znamená colocation analysis, „ur“ sú prvé písmená z kategórie záujmu „urban fabric“ a zo susediacej kategórie „relax“, „ab“ označuje typ vstupnej vrstvy, ktorá obsahuje len body kategórie záujmu (A) a susediacej kategórie (B) a žiadne ďalšie body iných kategórií (C). „800“ je dĺžka vzdialenosti v m definujúcej okolie, „g“ je prvé písmeno z váhovej schémy Gaussian a „499“ označuje počet permutácií.

Po nastavení všetkých potrebných parametrov je spustená kolokačná analýza a výstupná vrstva je pridaná do mapy. Užívateľ sa teda môže zoznámiť s výstupom analýzy v mape, cez atribútovú tabuľku a pomocou grafu, ktorý zobrazuje vzťah medzi LCLQ a p-value.

Priestorová interpretácia výsledkov

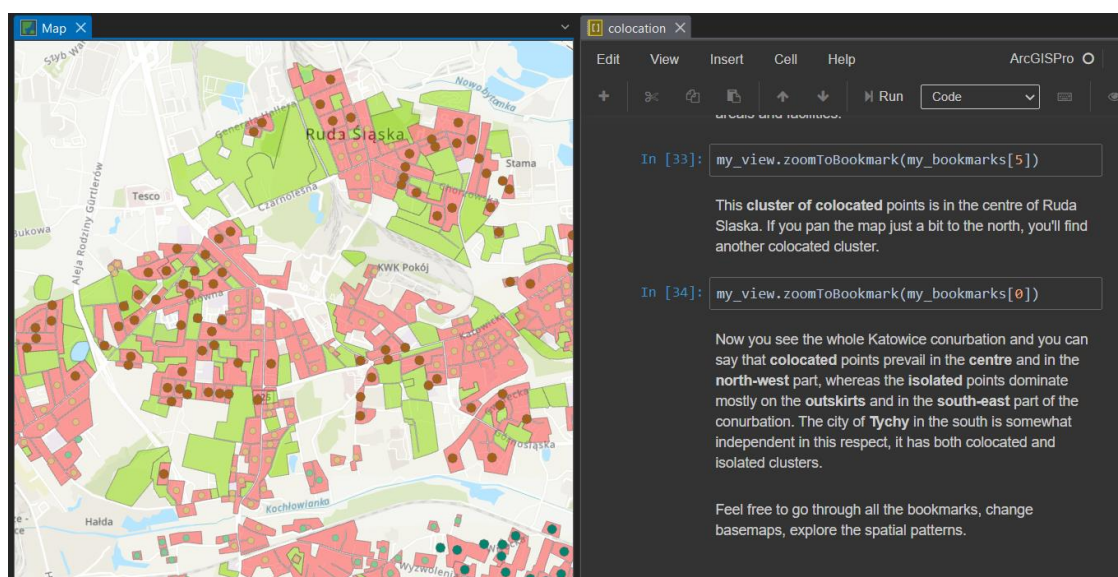
V tretej časti je užívateľ pomocou bookmarkov prevedený výstupnou vrstvou kolokačnej analýzy. Je to vrstva bodov kategórie záujmu a pre každý je vypočítaný LCLQ a p-value a na základe týchto dvoch hodnôt je každý bod priradený do jednej z 5 výsledných skupín (Obr. 20).

Symbol	LCLQ Bin	LCLQ Type	Description
	0	Colocated - Significant	Local colocation quotient is greater than 1 with a p-value less than 0.05.
	1	Colocated - Not Significant	Local colocation quotient is greater than 1 with a p-value greater than 0.05.
	2	Isolated - Significant	Local colocation quotient is equal to or less than 1 with a p-value less than 0.05.
	3	Isolated - Not Significant	Local colocation quotient is equal to or less than 1 with a p-value greater than 0.05.
	4	Undefined	The feature did not have any other features within its neighborhood or bandwidth equal to 0.

Obr. 20 Výsledné skupiny podľa LCLQ a p-value (zdroj: [How Colocation Analysis works](#)).

V tomto prípade to znamená, že kolokované body sú také body mestskej zástavby, ktoré majú do vzdialenosti 800 m väčší podiel bodov skupiny „relax“ (zeleň a športoviská) ako je ich globálny podiel v celej skúmanej oblasti, tým pádom LCLQ je väčší ako 1. Na druhej izolované body majú vo svojom okolí menší podiel bodov skupiny „relax“ ako je ich globálny podiel, čiže LCLQ je menší ako 1. Na podiel bodov susediacej kategórie v okolí každého bodu kategórie záujmu má vplyv zvolená lokálna váhová schéma.

Kolokované a izolované body majú niekedy tendenciu vytvárať zhluky. Je tomu tak aj v tomto prípade a vytvorené bookmarky prevedú užívateľa práve takýmito zhlukmi. Najprv sú vrstvy v mape nastavené tak, aby sa zhluky bodov dali dobre interpretovať a potom pomocou MapView metódy [Zoom To Bookmark](#) užívateľ „cestuje“ v mape a v notebooku je k tomu napísaná jednoduchá interpretácia. Ako prvými prechádza 3 zhlukmi izolovaných bodov na juhu Katowíc, v Mysłowiciach a v Javorzne. Nasledujú 3 zhluky kolokovaných bodov v mestách Sosnowiec, Ruda Śląska (Obr. 21) a Gliwice.



Obr. 21 Ukážka zhluku kolokovaných bodov v centre mesta Ruda Śląska (autor: A. Tóth).

Posledný bookmark zobrazí celú katowickú konurbáciu a pri tomto pohľade je vidieť, že kolokované body prevažujú v centre a severozápadnej časti, zatiaľ čo izolované body sa sústreďujú viac na okrajoch a v juhovýchodnej časti. Môže to byť spôsobené tým, že v centre býva viac parkov, lavičiek, ihrísk pre deti a športových priestorov než na okrajoch miest.

Spustenie viacerých analýz

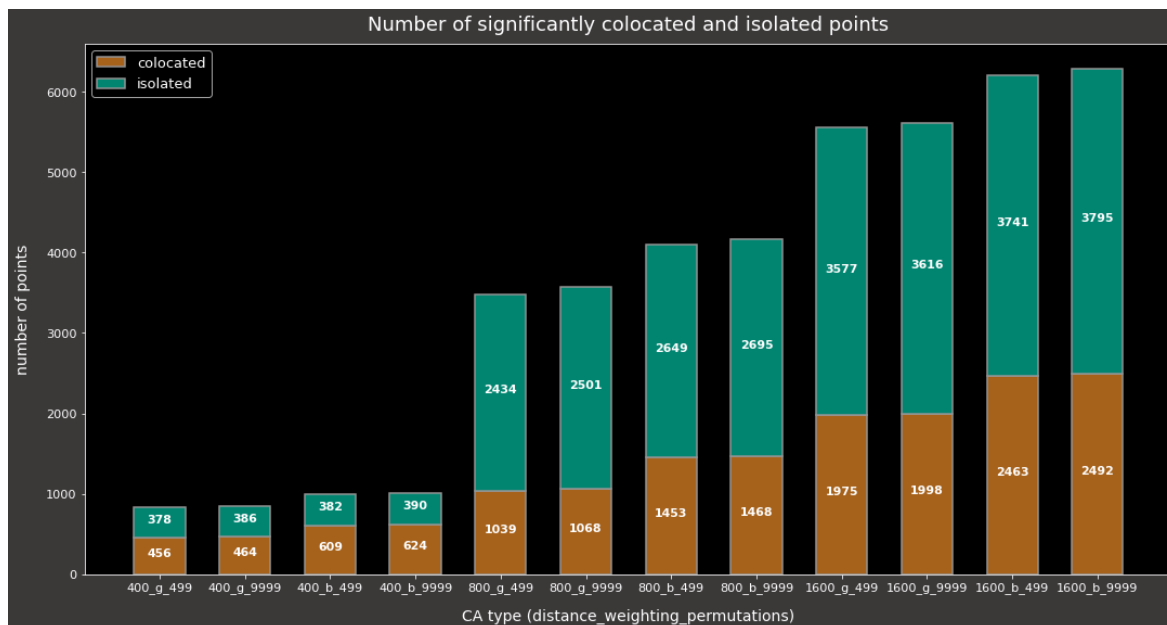
Štvrtá časť využíva for cykly na spustenie viacerých analýz naraz s rôznymi nastaveniami. Parametre ako vstupná vrstva, kategórie A a B a typ okolia ostávajú rovnaké. Mení sa však veľkosť okolia, pôvodných 800 m sa vydolí a vynásobí 2 a analýzy budú spustené so 400, 800 a 1600 m vzdialenosťou určujúcou veľkosť okolia. Meniť sa bude aj počet permutácií, testovať sa bude stredný počet 499 a najvyšší počet 9999 a tiež budú 2 typy lokálnej váhovej schémy, Gaussian a Bisquare. Celkovo bude teda kolokačná analýza spustená 12-krát, keďže $3 \times 2 \times 2 = 12$ (3 vzdialenosti, 2 počty permutácií, 2 váhové schémy). Názvy výstupných vrstiev sa riadia rovnakými pravidlami/konvenciami ako pri prvej analýze a sú ukladané do zoznamu. Výstupné vrstvy sú automaticky pridávané do mapy a následne je z nich vytvorená skupina pre lepšiu manipuláciu.

Vizualizácia výsledkov analýz

Piata časť sa venuje vizualizácii a interpretácii výsledkov týchto 12 analýz. Najprv je vytvorená dátová štruktúra slovník, do ktorej sú uložené počty významne kolokovaných a izolovaných bodov. Z tohto slovníka je vytvorený [Pandas DataFrame](#), ktorý vytvorí prehľadnú tabuľku a ktorý je zároveň vstupom pre Matplotlib graf (Obr. 22). Je to tzv. „skladaný“ stĺpcový graf (R CODER 2023), kde každý stĺpec reprezentuje jednu z 12 analýz. Stĺpec je rozdelený na 2 časti, výška spodnej časti znázorňuje počet významne kolokovaných bodov a výška hornej časti zasa počet významne izolovaných bodov. Na ose y sú teda absolútne počty bodov, na ose x sú jednotlivé analýzy pomenované podľa

meniacich sa nastavení (vzdialenosť určujúca veľkosť okolia, lokálna váhová schéma, počet permutácií) v tvare 400/800/1600_g/b_499/9999.

Z grafu je vidno, že významne kolokovaných bodov je viac ako izolovaných iba pri vzdialenosti 400 m, pri vzdialenosti 800 a 1600 m je viac izolovaných bodov ako kolokovaných.



Obr. 22 Graf počtu významne kolokovaných a izolovaných bodov vo výstupoch 12 analýz s rôznymi nastaveniami (autor: A. Tóth).

Výpočet kolokačného koeficientu

Posledná šiesta časť sa detailne venuje procesu výpočtu LCLQ bez lokálnej váhovej schémy. Pôvodným zámerom bolo zamerať sa na výpočet s Gaussian váhovou schémou, ale pre nedostatok zdrojov, ktoré by vysvetľovali jednotlivé premenné vo vzorci tejto váhovej schémy a pre jednoduchšie pochopenie princípu bolo nakoniec pristúpené k výpočtu bez váhovej schémy.

Na začiatku sa spustí kolokačná analýza s rovnakými nastaveniami aké mala prvá analýza, jediným rozdielom je nastavenie None v parametri lokálnej váhovej schémy. Potom je z výstupnej vrstvy tejto analýzy vybraný bod, pre ktorý sa bude „ručne“ počítať LCLQ podľa vzorca na Obr. 23.

$$LCLQ_{Ai \rightarrow B} = \frac{N_{Ai \rightarrow B}}{N_B / (N - 1)}$$

Obr. 23 Vzorec na výpočet LCLQ (zdroj: [How Colocation Analysis works](#)).

$LCLQ_{Ai \rightarrow B}$ je lokálny kolokačný koeficient i-tého bodu A pre susediace body B. $N_{Ai \rightarrow B}$ je podiel susediacich bodov B v okolí i-tého bodu A, čiže počet susediacich bodov B vydelený celkovým počtom susediacich bodov. N_B je počet všetkých bodov B v skúmanej oblasti

a N je počet všetkých bodov v skúmanej oblasti. Od tohto N je odpočítaný 1 bod preto, lebo samotný i -tý bod A nemôže byť zahrnutý v tomto počte a preto musí byť odrátaný.

Vybraný bod A a všetky jeho susediace body sú exportované do samostatných vrstiev, aby si mohol užívateľ lepšie predstaviť proces výpočtu a sám overiť napr. podiel bodov B v okolí. Po získaní všetkých potrebných premenných pomocou ArcPy funkcií `GetCount`, `SelectLayerByLocation` a `SelectLayerByAttribute` je vypočítaný LCLQ pre vybraný bod a na potvrdenie správnosti je porovnaný s LCLQ vypočítaným nástrojom `Colocation analysis`.

4.3 Hustota dopravnej siete

Umiestnenie a názov súboru: `/2_network_density/network_density.ipynb`

Vstupné dáta: OSM, GEOSTAT 2018 1 km² populačný grid

Konkrétny prípad: hustota železničnej siete mesta Liverpool v hexagónovej mriežke (rozloha hexagónu = 2 km²)

Potrebný komerčný software: áno, ArcGIS Pro

Počet markdown buniek: 64

Počet buniek kódu: 44

Počet riadkov kódu: 247

Úvod

Notebook vychádza zo skriptu [eu_grid_population.py](#), ktorý je jedným z výstupov bakalárskej práce Možnosti výpočtu charakteristik dopravní sítě států a měst (Tóth 2022). Skript je súčasťou ArcGIS toolboxu a rozhranie je typické ako pre akýkoľvek geoprocessingový nástroj. Užívateľ nastaví parametre, vloží vstupné dáta a spustí skript. Výsledkom je vrstva s hexagónovou mriežkou, ktorá obsahuje atribúty s vypočítanou hustotou dopravnej siete na plochu a na počet obyvateľov.

Užívateľovi je v tomto notebooku predstavený postup výpočtu, skript je rozdelený na menšie časti a je k nemu napísaný vysvetľujúci text. Veľkou výhodou oproti pôvodnému skriptu je stiahnutie dát priamo v notebooku, pretože pred spustením skriptu si musel užívateľ stiahnuť potrebné dáta sám.

Tento notebook nie je šírený spolu s ArcGIS projektom, užívateľ si projekt založí sám a vloží si doňho notebook. Kernelom je naklonované ArcGIS Python prostredie, do ktorého je potrebné doinštalovať tieto knižnice: `Folium` (Story 2024), `OSMnx` (Boeing 2017), `GeoPandas` (GeoPandas developers 2022), `Requests` (Reitz 2019) a `zipfile` (Python Software Foundation 2024c). Notebook bol použitý a otestovaný vo výuke na predmete KGI/PRODA v letnom semestri v roku 2024.

Dopravná sieť

Dopravná sieť je pojem, ktorý zahŕňa všetku dopravnú infraštruktúru: cesty, železnice, prístavy, námorné trasy, letiská, letecké trasy, potrubia, križovatky, terminály a pod. Tento notebook je zameraný na líniové prvky dopravnej siete, keďže počíta hustotu ako dĺžku siete na rozlohu územia alebo na počet obyvateľov na danom území. Nezáleží na konkrétnom dátovom zdroji, jediná požiadavka je na geometrický typ: vstupné dáta musia byť línie. V tomto prípade je zdrojom dát OSM, odkiaľ sú získané železnice mesta Liverpool.

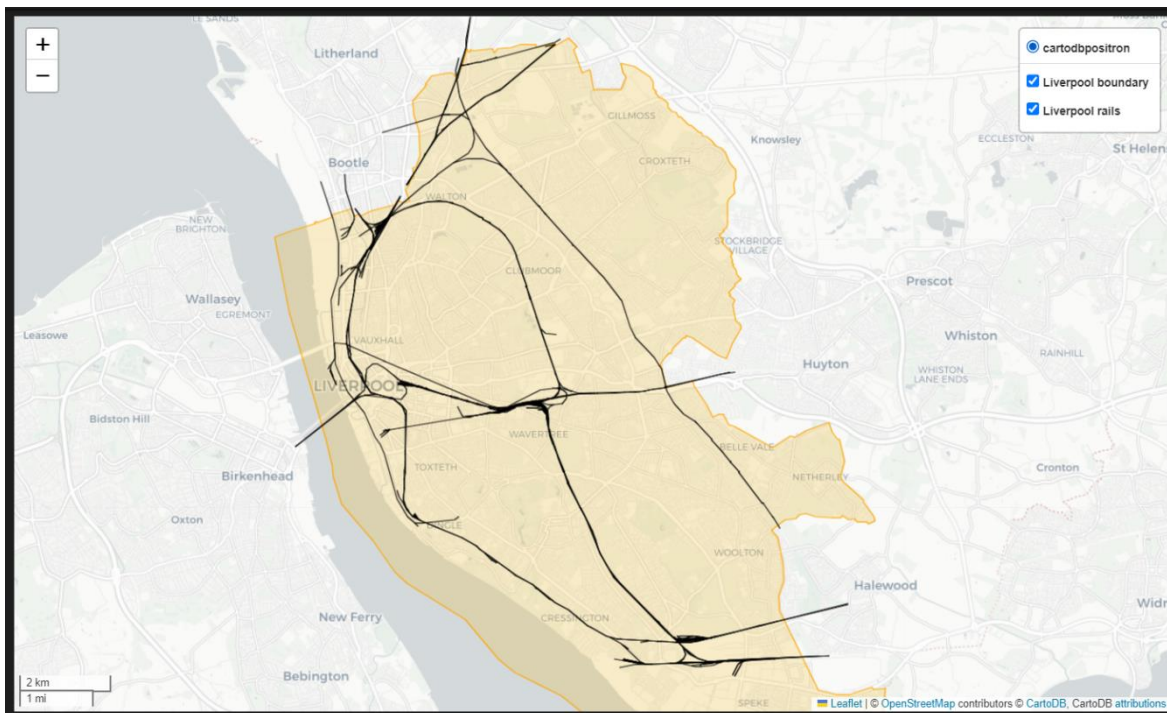
ArcGIS nastavenia

V prvej časti prebehnú štandardné nastavenia ArcGIS prostredia: importuje sa knižnica ArcPy, užívateľ nastaví cestu ku geodatabáze jeho ArcGIS projektu ako pracovný priestor, a povolí sa prepisovanie výstupov s rovnakým názvom.

Dáta o dopravnej sieti

Nasleduje získanie polygónu ohraničujúceho mesto Liverpool a jeho železničnej siete z OSM. Pomocou knižnice OSMnx a jej funkcie [geocode to gdf](#) sa do formátu GeoDataFrame uloží polygón ohraničujúci mesto Liverpool a je pomenovaný „liverpool_boundary“. Pre lepšiu predstavu, ako tento polygón vyzerá, je vizualizovaný v Leaflet mape vytvorenej knižnicou Folium. Následne je uložený vo formáte GeoJSON do adresára, kde sa nachádza aj ArcGIS projekt. Je to medzikrok potrebný na to, aby mohla byť vrstva „liverpool_boundary“ importovaná do geodatabázy v ArcGIS projekte pomocou nástroja [JSON To Features](#).

OSMnx umožňuje pristupovať k dátam OSM na vybranom mieste použitím funkcie [features from place](#). Jedným z parametrov tejto funkcie je výber vrstiev pomocou špecifikovania ich OSM tagu. Týmto spôsobom sú vybrané železnice (tag = {railway': True}) v Liverpoole a uložené do formátu GeoDataFrame. OSM dáta obsahujú veľké množstvo rôznych atribútov. Pre výpočet hustoty siete je však potrebná iba dĺžka línií, preto sú všetky atribúty okrem atribútu s geometriou odstránené. Tagom railway sú označené nielen líniové prvky, takže dáta v GeoDataFrame sú rôznych geometrických typov. Filtráciou podľa geometrického typu sú teda ponechané len líniové prvky. Finálna vrstva železníc spolu s polygónom „liverpool_boundary“ sú zobrazené na jednoduchej mape vytvorenej knižnicou Folium (Obr. 24).



Obr. 24 Jednoduchá vizualizácia stiahnutých železníc a polygónu „liverpool_boundary“ (autor: A. Tóth).

V tejto fáze sú už železnice pripravené na vstup do ArcGIS prostredia na výpočet hustoty a preto je aplikovaný rovnaký postup ako pri polygónovej vrstve

„liverpool_boundary“ – železnice sú exportované do formátu GeoJSON a importované do geodatabázy v ArcGIS projekte nástrojom JSON To Features.

Dáta o populácii

Údaje o počte obyvateľov, ktoré budú vstupovať do výpočtu hustoty siete, pochádzajú z GEOSTAT populačného gridu o veľkosti štvorca 1 km². Tento grid pokrýva osídlené oblasti Európy a každý štvorec obsahuje počet obyvateľov, ktorý na danom území žijú. V tomto notebooku je stiahnutý pomocou knižnice Requests, ktorá umožňuje posielat http požiadavky a ukladať odpovede na ne. Je potrebné zadať URL adresu, na ktorej sa nachádza populačný grid, a tiež cestu k adresáru, kam sa má na disku uložiť. Je odporúčané uložiť stiahnutý grid do adresára, kde sa nachádza príslušný ArcGIS projekt a už sú tam uložené vrstvy železníc a „liverpool_boundary“. Ak je príslušná bunka s kódom spustená dva a viac krát a populačný grid je už stiahnutý, nebude sa sťahovať znova. Táto podmienka sa nachádza v notebooku vo všetkých bunkách, ktoré vytvárajú novú vrstvu alebo nový atribút. Vždy sa najskôr skontroluje, či vytváraný objekt už náhodou neexistuje až potom je vytvorený.

Ďalším krokom je rozbalenie zazipovaného súboru, ktorý obsahuje populačný grid v rastrovom formáte tif a vo vektorovom formáte shp. Použitá je funkcia [extract_all](#) z knižnice zipfile, ktorá vie prečítať a rozbaľiť zazipované súbory.

Do geodatabázy ArcGIS projektu nie je importovaný celý grid kvôli jeho veľkosti (313 MB vo formáte shp) a preto prebehne prvé „orezanie“ ešte mimo geodatabázy. Pomocou nástroja [Select Layer By Location](#) sú vybrané iba tie štvorce gridu, ktoré sa pretínajú akýmkoľvek spôsobom s polygónom „liverpool_boundary“. Vybrané štvorce sú skopírované do geodatabázy funkciou [Copy Features](#).

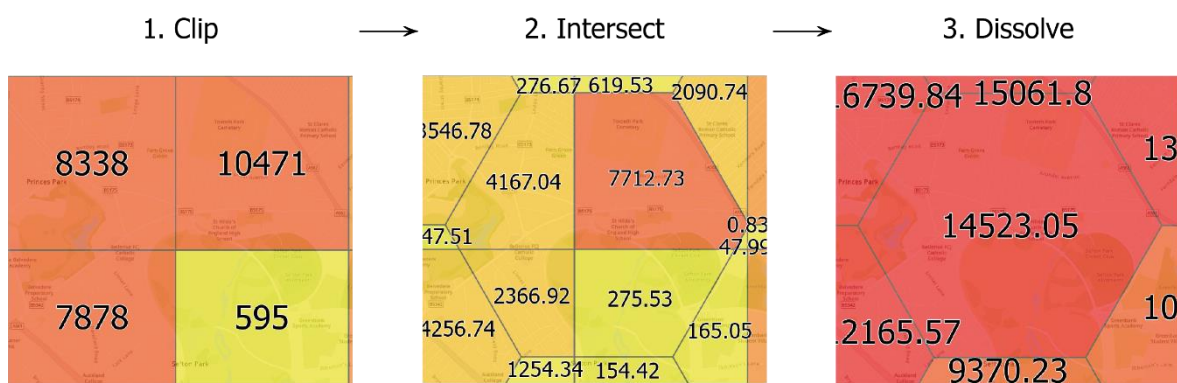
Nasledujúca činnosť bude vyžadovať zjednotené súradnicové systémy všetkých vstupných vrstiev na dosiahnutie správnych a presných výsledkov. Odlišné súradnicové systémy populačného gridu, železníc a hexagónovej mriežky môžu spôsobiť posunutie týchto vrstiev tak, že sa nebudú prekrývať a tým pádom nebude môcť byť spočítaná hustota železničnej siete. Preto je v tomto bode kľúčové nastavenie súradnicového systému, do ktorého budú projektované všetky výstupy geoprocessingových nástrojov. Za tento jednotný súradnicový systém je zvolený ETRS 1989 LAEA (epsg: 3035), v ktorom je uložený populačný grid. Pomocou funkcie Describe a atribútu [spatial reference](#) sa k nemu pristúpi a nastaví sa ako [arcpy.env.outputCoordinateSystem](#).

Nasleduje úprava pred orezaním populačného gridu polygónom „liverpool_boundary“. Tie štvorce z gridu, ktoré ležia na hranici polygónu, neostanú po orezaní celé, ostane iba ich časť. Avšak počet obyvateľov v atribútovej tabuľke v orezanom štvorci sa nezmení. Preto je potrebné spraviť prepočet populácie. Keďže podrobnejšie dáta, z ktorých vychádza GEOSTAT populačný grid, nie sú k dispozícii, prepočíta sa počet obyvateľov v nekompletnom štvorci proporcionálne podľa rozlohy. Tento prepočet predpokladá rovnomerné rozdelenie populácie na celom území štvorca. Do atribútovej tabuľky gridu je pridaný nový stĺpec, do ktorého je pomocou Calculate Field uložená pôvodná rozloha každého štvorca. Následne prebehne orezanie nástrojom Clip a prepočet populácie, ktorý je realizovaný pridaním ďalšieho nového stĺpca, do ktorého je napočítaný aktuálny počet obyvateľov podľa vzorca pôvodný počet obyvateľov vydelený pôvodnou rozlohou a vynásobený aktuálnou rozlohou.

Hexagónová mriežka

V tejto časti je najprv vygenerovaná hexagónová mriežka s veľkosťou hexagónu 2 km² nástrojom [Generate Tessellation](#) a následne je orezaná polygónom „liverpool_boundary“. Ďalším krokom bude vytiahnutie informácií o dĺžke železníc a o počte obyvateľov pre každý hexagón.

V prvom rade to je počet obyvateľov. Do stĺpca orezaného populačného gridu sa uloží aktuálna rozloha a grid je rozsekaný nástrojom Intersect podľa hexagónovej mriežky. Vo vrstve s rozsekanými štvorcami populačného gridu prebehne rovnaký prepočet počtu obyvateľov, aký prebehol po orezaní gridu polygónom „liverpool_boundary“ (pôvodná populácia delená pôvodnou rozlohou a násobená aktuálnou rozlohou). Nástrojom Dissolve je realizované zjednotenie rozsekaných častí populačného gridu podľa toho, do ktorého hexagónu patria. Spomínaný nástroj ponúka rôzne možnosti, ako manipulovať s jednotlivými atribútmi vstupnej vrstvy, medzi ktorými je aj sčítanie hodnôt. Práve táto možnosť je vybraná pre atribút s prepočítanou populáciou. Pomocou funkcie Join Field je atribút so sčítaným počtom obyvateľov pre každý hexagón pripojený k vrstve hexagónovej mriežky. Schéma celého procesu je zachytená na Obr. 25.



Obr. 25 Ilustrácia získania počtu obyvateľov z populačného gridu do hexagónovej mriežky (autor: Adam Tóth).

Keďže populačný grid pokrýva iba osídlené územia a polygón „liverpool_boundary“ zahŕňa aj neosídlené časti, ktoré sa nachádzajú na území mesta Liverpool, niektoré hexagóny budú obsahovať nulový počet obyvateľov, čo sa v pripojenom atribúte s počtom obyvateľov prejaví hodnotou NULL. Túto hodnotu je nutné prepísať na číselnú hodnotu 0, pretože počet obyvateľov bude pri výpočte hustoty menovateľom a ak by obsahoval hodnoty NULL, viedlo by to k zlyhaniu výpočtu. Na toto prepísanie je použitá funkcia Calculate Field.

Podobný postup je uplatnený pri získaní dĺžky železníc. Dáta z OSM sú pôvodne stiahnuté v súradnicovom systéme WGS 1984 (epsg: 4326) a preto musí byť vrstva železníc najprv projektovaná nástrojom [Project](#) do jednotného súradnicového systému, ktorým je ETRS 1989 LAEA (epsg: 3035). Následne sú železnice rozsekané a zjednotené podľa hexagónovej mriežky nástrojmi Intersect a Dissolve. Do nového stĺpca je uložená dĺžka železníc pre každý hexagón. Tento stĺpec je funkciou Join Field pripojený do vrstvy hexagónovej mriežky. Nie všetky hexagóny majú na svojom území železničnú trať a preto niektoré z nich majú v pripojenom stĺpci s dĺžkou železníc hodnotu NULL. Hoci pre výpočet hustoty nie je potrebné prepisovať NULL na 0 (dĺžka železníc bude čitateľom), pre udržanie konzistentnosti dát je to tak spravené rovnako ako pri stĺpci s počtom obyvateľov.

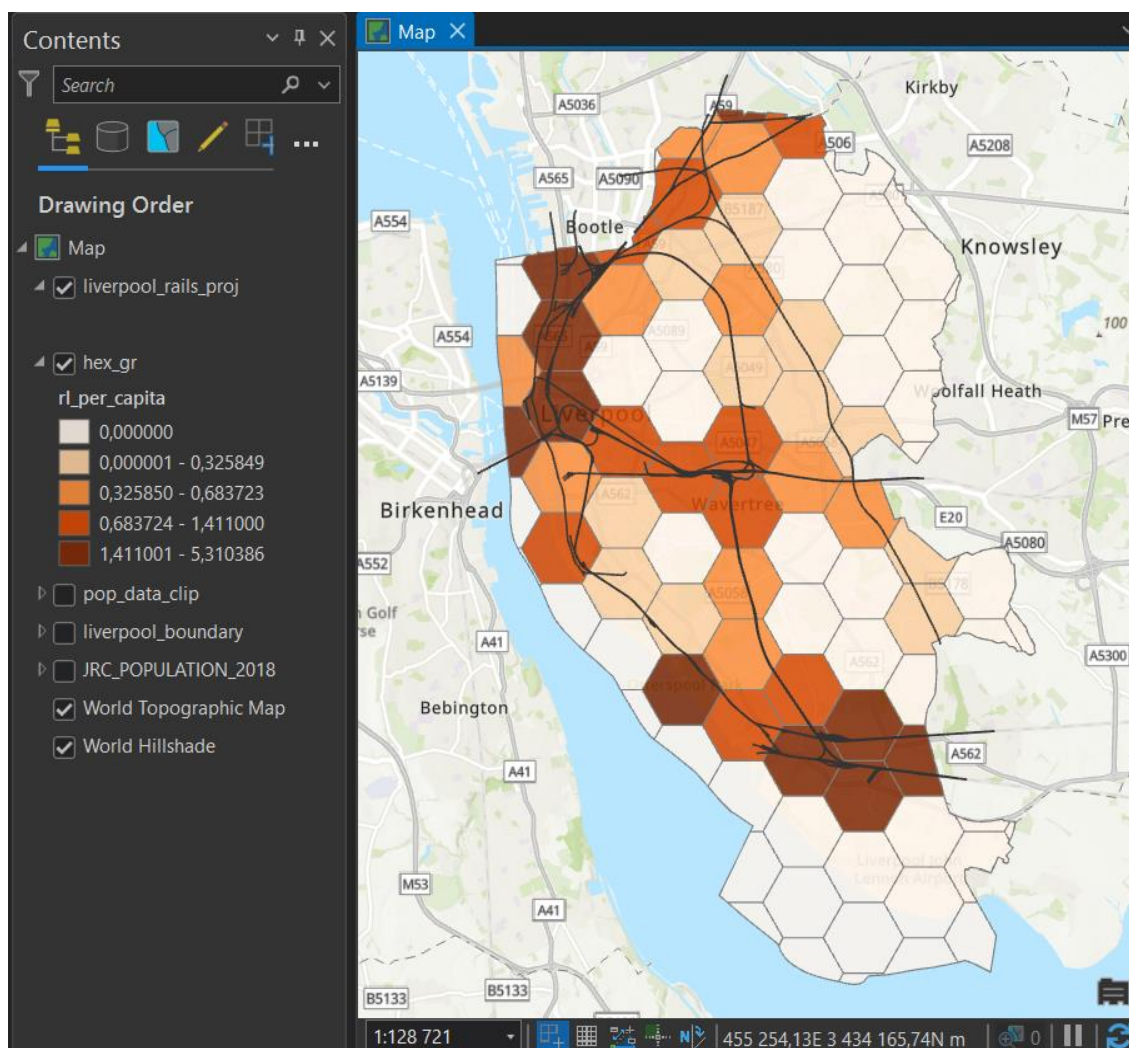
Výpočet hustoty

V tejto fáze vrstva hexagónovej mriežky obsahuje všetky potrebné informácie na výpočet oboch typov hustôt, pre ktoré sú vytvorené nové stĺpce funkciou [Add Fields](#). Ako prvá je vypočítaná hustota železníc na plochu ako dĺžka železníc v km na 1 km² rozlohy na ktorej sa nachádzajú. Použitý je nástroj Calculate Field.

Ako druhá je vypočítaná hustota železníc na populáciu ako dĺžka železníc v m na 1 obyvateľa. Na tento výpočet je použitá funkcia [Update Cursor](#), ktorá vytvorí objekt, cez ktorý sa iteruje vo for cykle a prístupuje sa k jednotlivým hexagónom. Tento prístup je použitý z dôvodu, že niektoré hexagóny majú v atribúte s počtom obyvateľov hodnotu 0 a delenie 0 spôsobuje pád výpočtu. Preto sa hustota počíta iba pre hexagóny, ktoré sú osídlené. Neosídleným hexagónom bude v atribúte hustota na populáciu priradená hodnota 0. Na záver sú z geodatabázy vymazané pomocné a nepotrebné vrstvy funkciou [Delete](#).

Vizualizácia

Napočítané hustoty je možné vizualizovať pomocou gradovaných farieb (Obr. 26) alebo gradovaných symbolov.



Obr. 26 Ukážka vizualizácie hustoty železničnej siete na počet obyvateľov (autor: A. Tóth).

4.4 Eurostat

Umiestnenie a názov súboru: `/3_eurostat/eurostat.ipynb`

Vstupné dáta: Eurostat

Konkrétny prípad: dataset Human resources in science and technology (HRST) by NUTS 2 regions, kód `tgs00038`

Potrebný komerčný software: nie

Počet markdown buniek: 75

Počet buniek kódu: 57

Počet riadkov kódu: 334

Úvod

Ako bolo už naznačené v predchádzajúcom notebooku o hustote dopravnej siete, veľkou výhodou Pythonu je sťahovanie a prístup k dátam pomocou rôznych API a knižníc. ArcGIS (a vo všeobecnosti Jupyter) notebook má možnosť hneď zobrazíť stiahnuté dáta, či už vo forme tabuľky, grafy alebo mapy. V tomto notebooku je dôraz kladený práve na vyššie spomenuté činnosti – jednoduchý prístup k dátam cez knižnice Eurostat (Cazzaniga 2024) a Requests a ich následné spracovanie a vizualizácia pomocou knižníc Pandas, GeoPandas, Matplotlib a matplotlib-scalebar (Pinard 2022). Užívateľ si musí nainštalovať týchto 6 knižníc do svojho Python prostredia, ktoré mu bude slúžiť ako kernel pre tento notebook. Pri vývoji a testovaní notebooku bolo použité naklonované ArcGIS Python prostredie 3.9.18., avšak knižnica ArcPy ani iná funkcionálna ArcGISu nie je potrebná a užívateľ môže použiť akékoľvek Python prostredie ako kernel na spúšťanie kódu v tomto notebooku. Tiež výber programu, v ktorom bude notebook používaný a spúšťaný, je na preferencii užívateľa, môže to byť ArcGIS Pro, VS Code, JupyterLab, Google Colab a ďalšie prostredia, ktoré vedú pracovať s Jupyter Notebookom. Odporúčanie autora je program VS Code.

Eurostat (databáza, GISCO, API)

Ako každá krajina má svoj štatistický úrad, tak aj EÚ má svoj a je ním Eurostat. Zhrmažďuje dáta členských krajín EÚ a ďalších európskych krajín, ktoré boli kedysi v EÚ (Spojené kráľovstvo), majú s EÚ blízke hospodárske vzťahy a sú súčasťou Schengenského priestoru (Island, Nórsko, Švajčiarsko) alebo sú kandidátmi na vstup do EÚ (Albánsko, Čierna Hora, Severné Macedónsko, Turecko, Srbsko). Databáza na webovej stránke Eurostatu má príjemné užívateľské rozhranie, jednoducho sa v nej dajú vyhľadávať datasey pomocou témy, názvu alebo ID. Po otvorení konkrétneho datasetu je užívateľovi ponúknutý tabuľkový náhľad, kde je možné filtrovať, pridávať a odoberať atribúty a záznamy podľa dostupných kategórií. Ďalej si užívateľ môže zobrazíť základný graf a interaktívnu mapu.

Súčasťou Eurostatu je aj GISCO. Tu sú uložené geografické dáta ako napr. administratívne hranice európskych regiónov NUTS 0-3 a populačný grid (Eurostat no date). Datasets v hlavnej databáze Eurostatu neobsahujú geometrickú zložku a preto pri tvorbe máp musia byť prepojené s príslušnou vrstvou z GISCO.

K Eurostat datasetom je možné pristupovať cez Eurostat Python Package, ktorého aktuálna verzia je 1.1.0. Táto knižnica vie „čítať dáta a metadáta z Eurostatu a ukladať ich do Python štruktúr ako zoznam n-tíc alebo ako Pandas DataFrame“ (Cazzaniga 2024). Uloženie dát a vrstiev do premenných je výhodou oproti priamemu sťahovaniu dát na lokálny disk počítača, pretože takto zaberajú iba miesto v RAM (Random Access Memory)

pamäti a nie je potrebné ich zapisovať na disk. Po zatvorení notebooku sú sice z RAM vymazané, ale pri ďalšom otvorení sa veľmi jednoducho a rýchlo načítajú vďaka pár riadkom kódu. Blížši pohľad na knižnicu Eurostat prinesie prvá časť notebooku.

Výber datasetu z Eurostatu

Prvým krokom je importovanie knižnice Eurostat a zavolanie funkcie [get toc df](#), čím sa zoznam všetkých Eurostat datasetov uloží do štruktúry DataFrame. Tento zoznam (table of content) obsahuje názov, kód, typ, časové ohraničenie a poslednú úpravu datasetov. Funkciou [subset toc df](#) je možné filtrovať datasety podľa kľúčových slov. V tomto prípade je kľúčovým slovom „science“ a výsledkom je 6 datasetov, ktoré toto slovo obsahujú v názve. Po prieskume týchto datasetov a po krátkej konzultácii s dr. Karlom Macků bol vybraný dataset „Human resources in science and technology (HRST) by NUTS 2 regions“ s kódom *tgs00038*. Ľudské zdroje vo vede a technológii sú definované ako percentuálny podiel aktívnej populácie vo veku 15-74 rokov, ktorá buď úspešne dokončila terciárne, čiže vysokoškolské vzdelanie alebo je zamestnaná vo vedeckom a technologickom sektore (Eurostat 2024). Inými slovami dáta ukazujú, aká časť pracujúcej populácie má potenciál pracovať vo vede a v technológii alebo už v tomto sektore pracuje. Dáta sú aktualizované s ročnou frekvenciou a aktuálne sú dostupné pre roky 2011-2022. Podrobnejšie informácie o tomto datasete je možné nájsť v jeho metadátovom zázname „Human Resources in Science & Technology (hrst): Reference Metadata in Euro SDMX Metadata Structure (ESMS)“ (Eurostat no date).

Dataset je pomocou ID vložený do funkcie [get data df](#) uložený do štruktúry DataFrame a zobrazený vo forme tabuľky hneď pod bunkou kódu. Funkciou [get pars](#) sú načítané názvy stĺpcov tohto datasetu a funkcia [get dic](#) vráti slovník s názvom stĺpca používaným v datasete, plným znením názvu a s významom stĺpca. Táto funkcia má viac parametrov a po chvíli čítania dokumentácie ich vie užívateľ nastaviť s použitím for cyklu tak, aby mu funkcia vrátila významy všetkých hodnôt v stĺpcoch datasetu. Týmto spôsobom užívateľ získa jasnejšiu predstavu o hodnotách a ich jednotkách, keďže väčšina je uvedená iba vo forme skratiek. Napríklad regióny NUTS 2 sú v datasete uvedené pod ich kódmi, resp. ID, takže získaním ich plných názvov je jasnejšie, o aké konkrétne regióny ide (Obr. 27).

```
for code, full_name in column_dict[3].items():
    if 'CZ' in code:
        print(f"{code}: {full_name}")
```

[10]

... CZ01: Praha
CZ02: Střední Čechy
CZ03: Jihozápad
CZ04: Severozápad
CZ05: Severovýchod
CZ06: Jihovýchod
CZ07: Střední Morava
CZ08: Moravskoslezsko

Obr. 27 Výpis všetkých regiónov NUTS 2, ktoré v ich kóde obsahujú písmená „CZ“ (autor: A. Tóth)

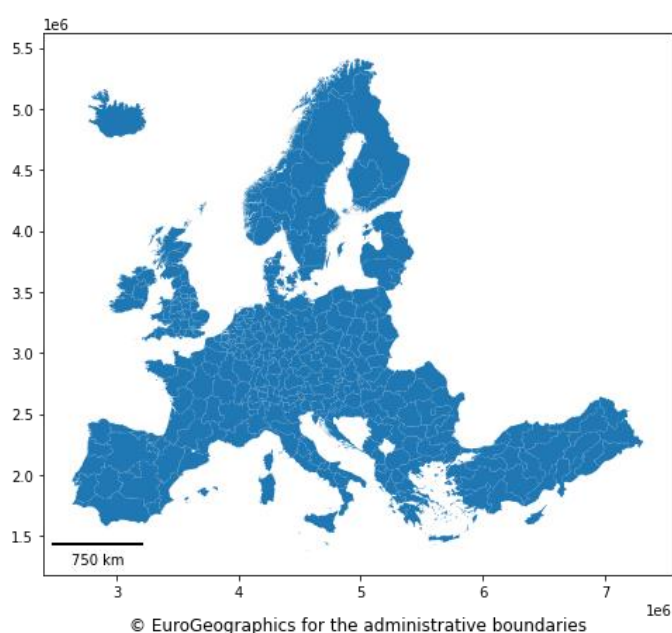
Načítanie NUTS 2 regiónov do formátu GeoDataFrame

Postupom podobným sťahovaniu populačného gridu v notebooku Hustota dopravnej siete je stiahnutý GeoJSON súbor s regiónmi NUTS 2. Použitím knižnice Requests je zaslaná http [get](#) požiadavka na URL [adresu](#) rozhrania GISCO services, kde sa tento GeoJSON nachádza. Pomocou knižnice os sa získa cesta k aktuálnemu adresáru (adresár, kde je uložený tento notebook), kam je tento GeoJSON súbor stiahnutý. Nakoniec je prečítaný funkciou [read file](#) z knižnice GeoPandas a regióny NUTS 2 sa uložia vo formáte GeoDataFrame. Ako už meno tejto štruktúry napovedá, je to vlastne Pandas DataFrame obohatený o geografickú zložku (priestorovú zložku, geometriu), takže jeden stĺpec je geometrického typu a obsahuje geografické súradnice.

Pri pohľade na prvých 5 riadkov tabuľky NUTS 2 regiónov vystúpia 3 a 2 stĺpce, ktoré môžu mať identické hodnoty: id, NUTS_ID, FID by mohli obsahovať tie isté kódy regiónov a NAME_LATN, NUTS_NAME by mohli obsahovať rovnaké názvy regiónov. Konverziou hodnôt týchto stĺpcov do zoznamov a ich následným porovnaním je zistené, že hodnoty v stĺpcoch s kódmi sú naozaj identické, zatiaľ čo názvy regiónov sú rôzne v spomínaných 2 stĺpcoch. Stĺpec NUTS_NAME obsahuje názvy v pôvodnom písme, takže regióny štátov, kde sa používa azbuka a cyrilika, sú napísané v azbuke a cyrilike. Stĺpec NAME_LATN obsahuje názvy všetkých regiónov v latinke. Po tejto kontrole sú duplicitné a nepotrebné stĺpce odstrániť. Zo stĺpcov s kódmi bol ponechaný atribút NUTS_ID.

Knižnica GeoPandas umožňuje vizualizovať GeoDataFrame jednoduchou statickou mapou vytvorenou funkciou [plot](#) založenou na knižnici Matplotlib. Vďaka tejto vizualizácii je vidieť, že v regiónoch NUTS 2 je zahrnutých viacero zámorských teritórií, ktoré často neobsahujú dáta a sú pre tento prípad nepodstatné. Odstrániť sa ich dá použitím súradnicového indexu [cx](#). V tomto prípade sa majú ponechať len regióny východne od 15° západnej geografickej dĺžky a medzi 34°-70° severnej geografickej šírky.

Poslednou úpravou je zmena súradnicového systému GeoDataFrame. Regióny boli stiahnuté v súradnicovom systéme WGS 1984 (epsg: 4326), v ktorom sa veľmi jednoducho robí vyššie uvedené filtrovanie pomocou súradnicového indexu. Avšak na ďalšiu prácu a tvorbu máp je oveľa vhodnejší súradnicový systém ETRS 1989 LAEA (epsg: 3035), ktorý zobrazuje celú Európu tak, že zachováva skutočné rozlohy štátov (Obr. 28). Prevod do zvoleného súradnicového systému je realizovaný funkciou [to crs](#).

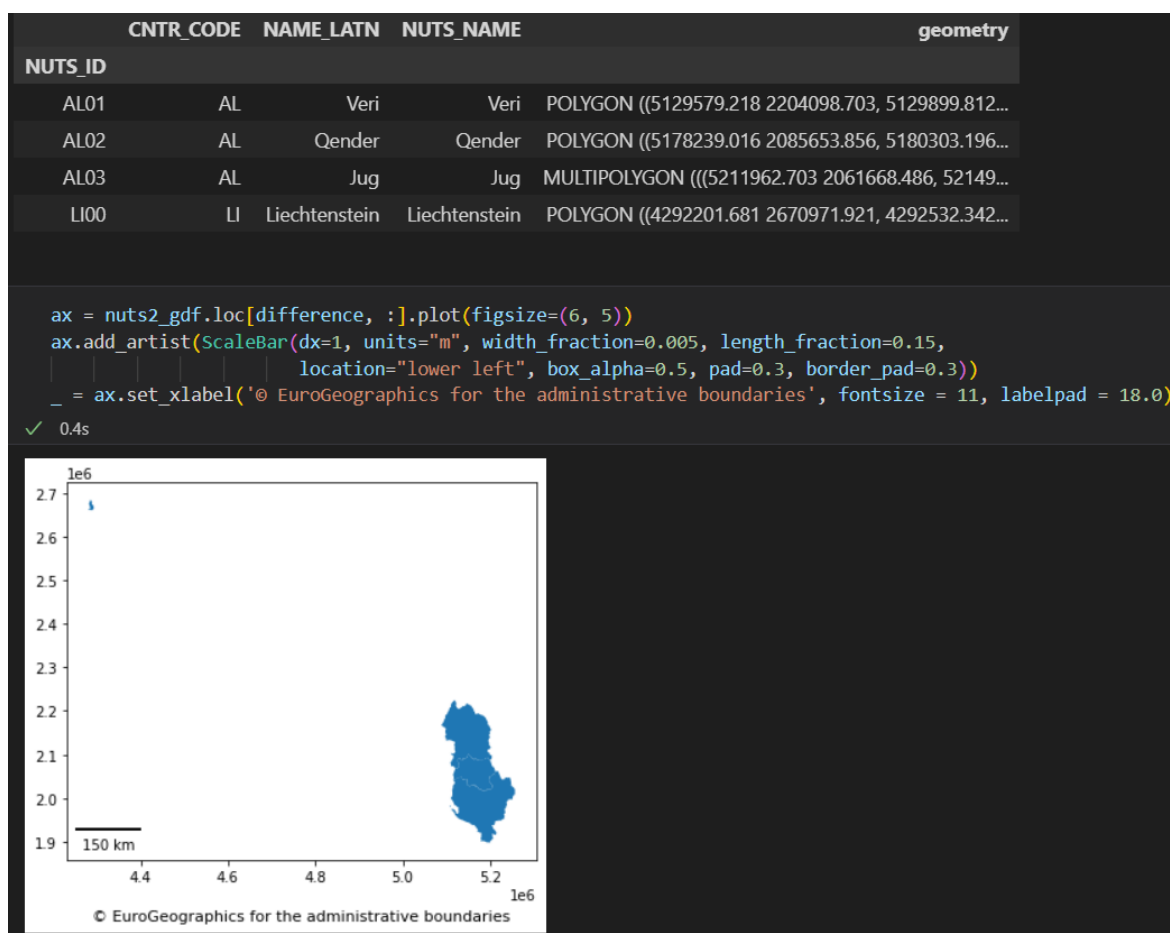


Obr. 28 Vizualizácia regiónov NUTS 2 po zmene súradnicového

Prepojenie geometrickej a atribútovej zložky

V tejto časti dochádza k prepojeniu Eurostat datasetu HRST (Human resources in science and technology) uloženom v DataFrame s vrstvou NUTS 2 regiónov uloženou v GeoDataFrame. Na prepojenie je potrebný spoločný atribút, čiže stĺpec s hodnotami, ktoré sa nachádzajú v oboch tabuľkách. V tomto prípade je to stĺpec s kódmi regiónov, v GeoDataFrame má názov NUTS_ID, v DataFrame zase geo. Stĺpec geo je teda premenovaný na NUTS_ID, aby prepojenie zavolané funkciou `merge` mohlo úspešne prebehnúť. Týmto spôsobom vzniká nový GeoDataFrame, ktorý obsahuje geometrickú zložku a zároveň dáta z datasetu HRST.

Výpisom počtu záznamov v pôvodnom a novovzniknutom GeoDataFrame je zistený rozdiel: v pôvodnej vrstve je o 4 záznamy viac ako v novej, obohatenej o dáta HRST. Je to z toho dôvodu, že pre 4 regióny neexistujú dáta HRST a preto boli pri prepojení „stratené“. Aby užívateľ zistil, o ktoré 4 regióny ide, potrebuje získať ich kódy, ktoré sa dajú získať porovnaním kódov v dvoch spomínaných vrstvách. Následne je možné ich zobraziť v tabuľke a v mape. Regióny NUTS 2, ktorým chýbajú dáta HRST, sú regióny Albánska a Lichtenštajnska (Obr. 29).



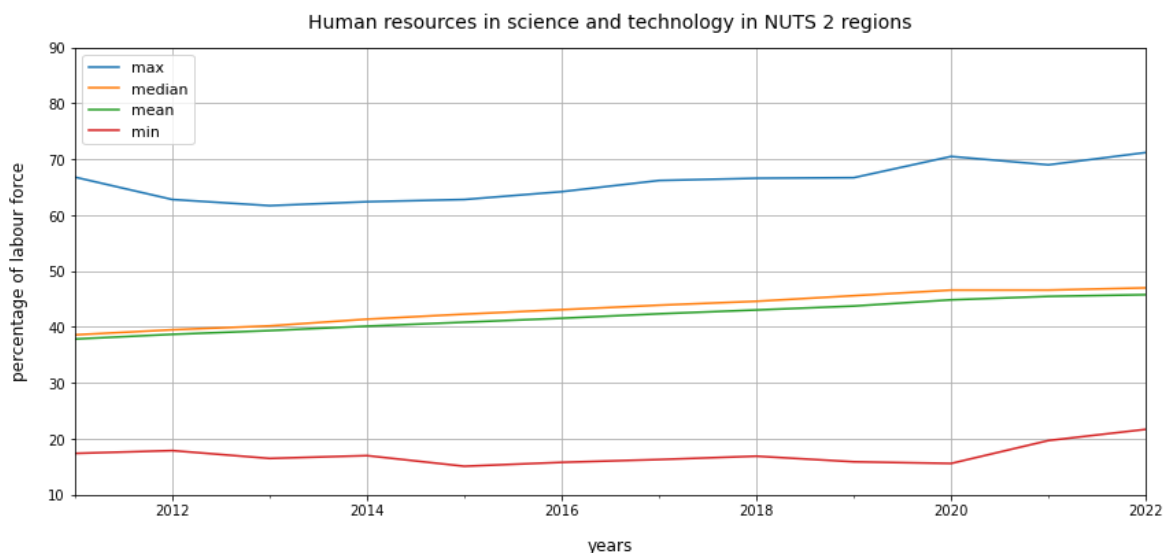
Obr. 29 Regióny NUTS 2, ktorým chýbajú dáta HRST (autor: A. Tóth).

Vizualizácie dát: grafy a mapy

Na začiatku tejto časti je importovaná knižnica Pandas, ktorá umožní lepšiu manipuláciu so štruktúrami (Geo)DataFrame. Funkcia `dropna` odstráni tie záznamy, ktoré majú null hodnoty. Sú to hlavne regióny Spojeného kráľovstva a štátov na Balkáne, keďže vzťah týchto krajín s EÚ sa menil v priebehu sledovaného obdobia a preto sú roky, kedy

dáta neboli dodávané. Vznikne tak nový GeoDataFrame, ktorý obsahuje len tie regióny, ktoré majú zaznamenaný percentuálny podiel HRST pre celé sledované obdobie (2011-2022) a bude vstupom pre nasledujúci graf, histogram a boxplot.

Prvým grafom bude líniový graf, ktorý bude zobrazovať celkový vývoj HRST v sledovanom období. Vývoj bude dokumentovaný minimálnou, maximálnou, priemernou a mediánovou hodnotou HRST v každom roku. Tieto hodnoty sú získané príslušnými funkciami z knižnice Pandas a uložené do samostatného DataFrame, kde riadkami sú roky a stĺpcami sú 4 sledované hodnoty. Z tohto DataFrame je vytvorený líniový graf (Obr. 30) rovnakou funkciou [plot](#), akou sa vytvára mapa z GeoDataFrame.

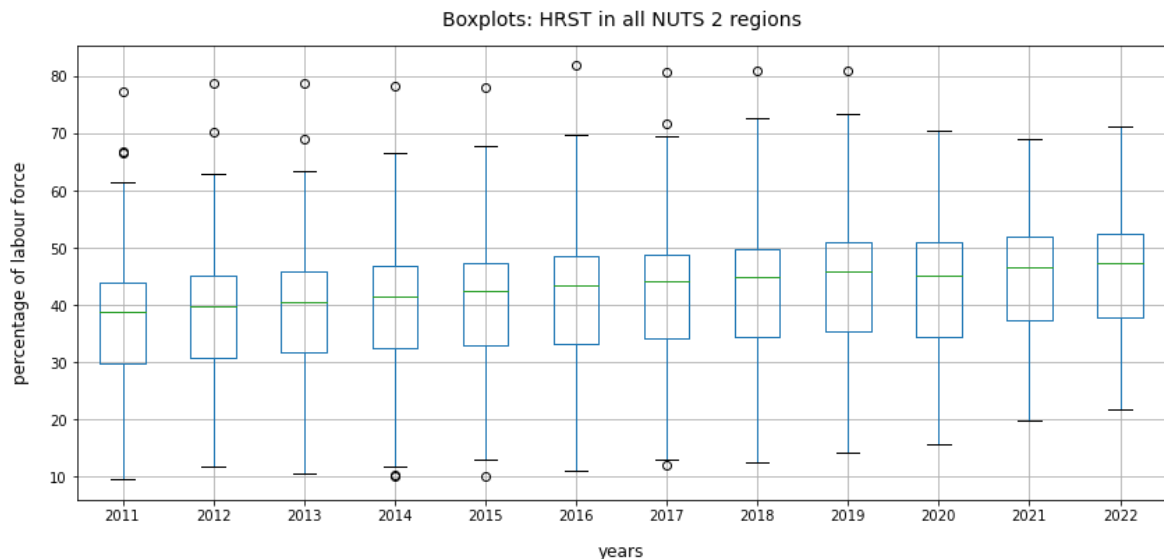


Obr. 30 Líniový graf zobrazujúci vývoj HRST v rokoch 2011-2022 (autor: A. Tóth).

Na grafe je možné pozorovať mierny nárast podielu aktívnej populácie s VŠ vzdelaním alebo zamestnaných vo vede a technológii, pričom medián sa celý čas drží tesne na priemerom.

Porovnanie rokov 2011 a 2022 je možné vidieť na histograme hodnôt HRST pre tieto dva roky. Histogramy potvrdzujú nárast zaznamenaný líniovým grafom, jedným z dôkazov je pozícia stĺpca reprezentujúceho najpočetnejšiu skupinu regiónov. Kým v roku 2011 najpočetnejšia skupina cca 60 regiónov mala hodnotu HRST okolo 40%, v roku 2022 sa táto skupina posunula na hodnotu HRST okolo 50%.

Boxploty vygenerované funkciou [boxplot](#) poskytujú prehľad o minime, maxime, mediáne, medzikvartilovom rozpätí a odľahlých hodnotách v každom roku. Odľahlých hodnôt je iba 5, po jednej v rokoch 2011, 2016, 2018, 2019, 2020. V ďalšom kroku sú vytvorené boxploty vychádzajúce z pôvodného GeoDataFrame, ktorý obsahuje aj regióny s null hodnotami. Tieto boxploty zobrazené na Obr. 31 sú „pestrejšie“ ako predchádzajúce, majú viac odľahlých hodnôt a aj medzikvartilové rozpätie sa mení vo väčšom rozsahu.



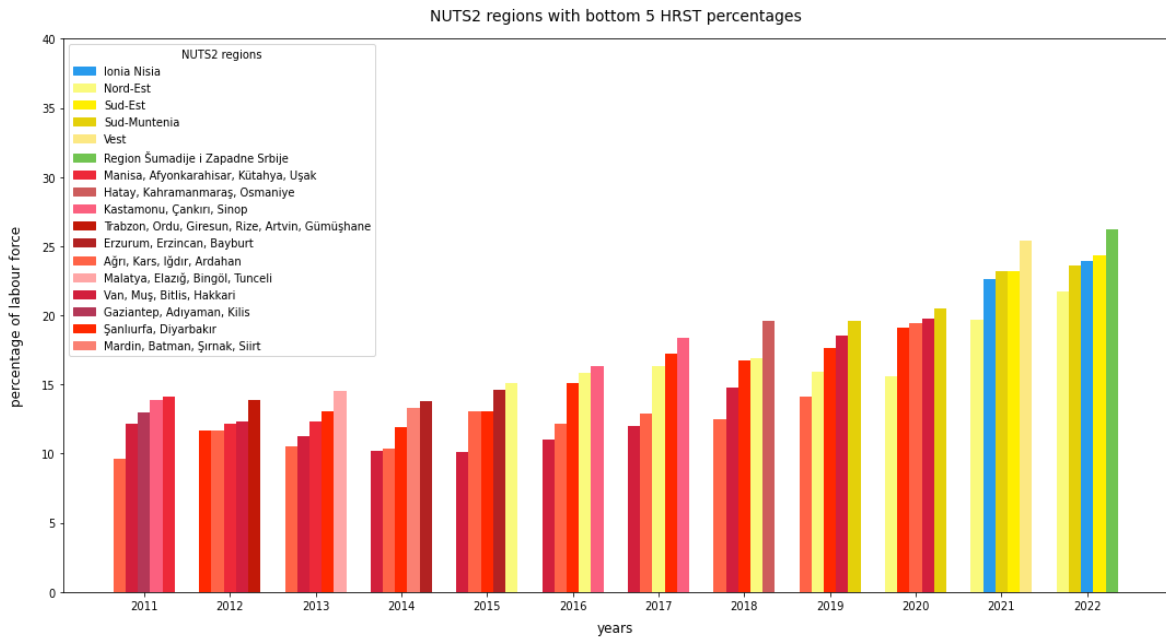
Obr. 31 Boxploty HRST hodnôt všetkých regiónov v rokoch 2011-2022 (autor: A. Tóth).

Nasledujúce bunky sa zameriavajú na výber a vizualizáciu regiónov, ktoré majú najnižší („najhoršie“ regióny), resp. najvyšší („najlepšie“ regióny) podiel HRST. Najprv je definovaná funkcia `tb5_selection`, ktorá vyberie 5 „najhorších“, resp. „najlepších“ regiónov v každom roku a uloží ich do `DataFrame`, ktorý vráti táto funkcia ako výstup. Jediným parametrom je boolean hodnota `True/False`, ktorá potom vstupuje do zoradenia HRST hodnôt a určuje, či sa zoradia vzostupne alebo zostupne. V prvom kole je funkcia zavolaná s parametrom `True`, ktorá pri tomto nastavení vráti 5 „najhorších“ regiónov vo všetkých rokoch. Výstupný `DataFrame` je uložený do premennej `bottom5_df`. Pandas funkciou [nunique](#) užívateľ zistí, že 17 regiónov sa strieda v najhoršej päťke počas celého sledovaného obdobia.

Druhá funkcia definovaná priamo v notebooku sa volá `get_unique_nuts` a má tiež iba jeden vstupný parameter, ktorým je `DataFrame` vytvorený predchádzajúcou funkciou `tb5_selection`. Funkcia `get_unique_nuts` vráti 2 zoznamy. Prvý obsahuje abecedne zoradené kódy regiónov NUTS 2, ktoré sa striedajú v najhoršej/najlepšej päťke počas celého sledovaného obdobia. Do druhého zoznamu sú uložené názvy týchto regiónov zoradené podľa ich abecedne zoradených kódov. Funkcia je zavolaná s parametrom `bottom5_df` a 2 výstupné zoznamy sú uložené do premenných `unique_b5_nuts` a `unique_b5_nuts_names`.

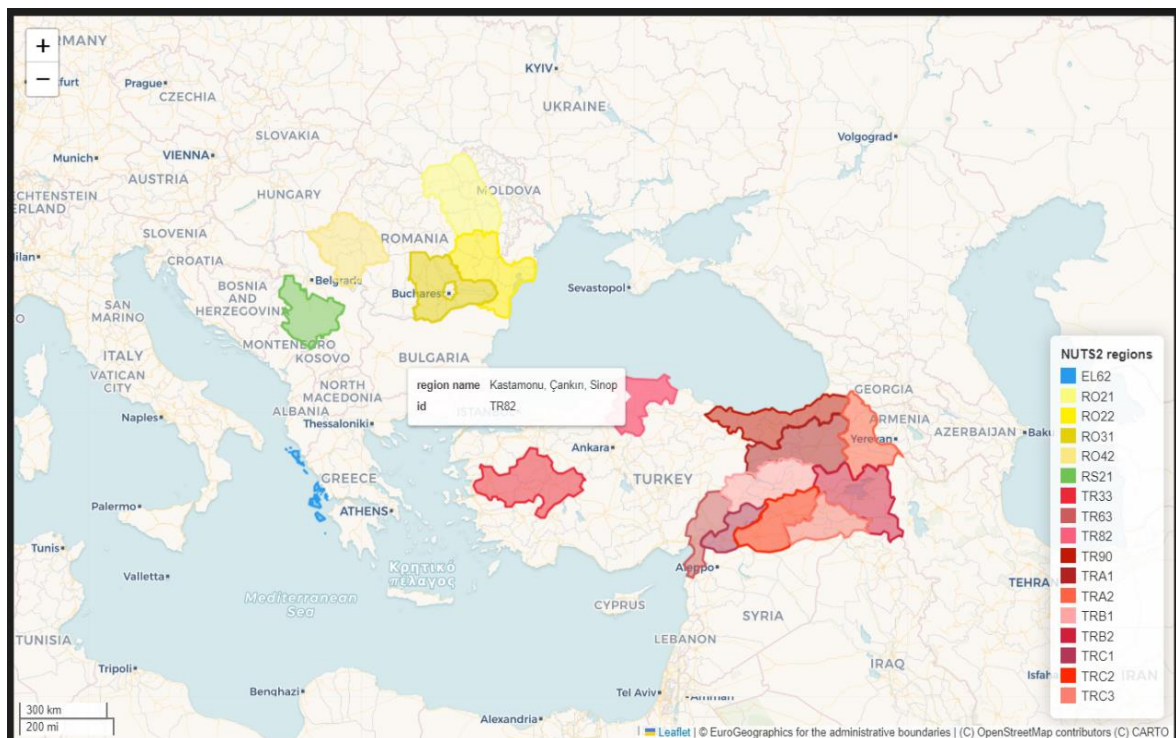
Po vypísaní kódov zo zoznamu `unique_b5_nuts` sú identifikované krajiny regiónov. V ďalšej bunke sú regiónom priradené farby podľa toho, z ktorej krajiny pochádzajú. Napríklad turecké regióny majú priradené rôzne odtiene červenej, rumunské regióny odtiene žltej atď. Tieto farby sú spolu s kódmi regiónov spojené metódou [zip](#) do slovníka `b5_color_dict`, kde kľúčom je kód a hodnotou je farba.

Najdlhšou a najzložitejšou funkciou definovanou v notebooku je funkcia `big_barplot`, ktorá má 7 parametrov. Táto funkcia pomocou submodulu [Pyplot](#) knižnice `Matplotlib` vytvorí stĺpcový graf zobrazujúci hodnoty HRST 5 najhorších, resp. najlepších regiónov v každom roku. Na osi x sú roky, na osi y sú hodnoty HRST ako percentuálny podiel pracujúcej populácie. Farby stĺpcov vychádzajú zo slovníka farieb. Ako prvý je vytvorený graf pre 5 najhorších regiónov (Obr. 32).



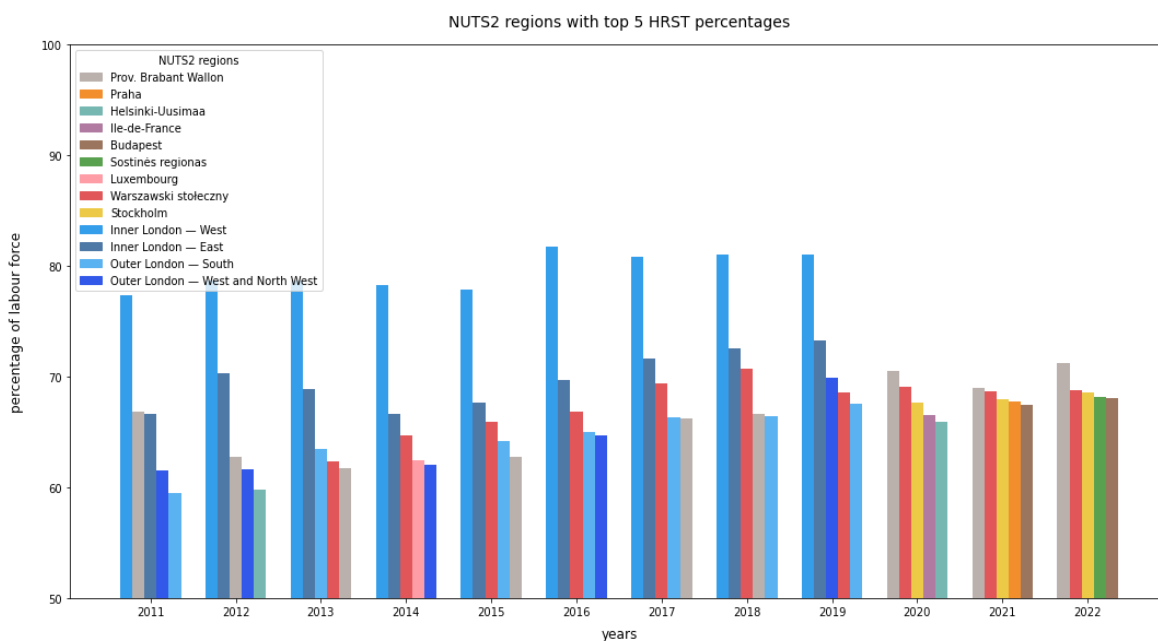
Obr. 32 Stĺpcový graf zobrazujúci 5 regiónov s najnižšími HRST podielmi (autor: A. Tóth).

Poslednou definovanou funkciou je funkcia `show_on_map`, ktorá pomocou funkcie [explore](#) vytvorí interaktívnu mapu 5 najhorších, resp. najlepších regiónov. Táto funkcia využíva Python knižnicu Folium, ktorá vytvára mapy založené na JavaScriptovej knižnici Leaflet.js. Vstupné parametre funkcie `show_on_map` sú zoznamy kódov, názvov a farieb regiónov. Na základe zoznamu kódov je vytvorený `GeoDataFrame`, pretože doteraz sa používala iba atribútová zložka a tak geometrická zložka nebola k vybraným regiónom pridaná. Po definícii funkcie je `show_on_map` zavolaná pre 5 najhorších regiónov. Pri prejdení myšou nad regiónom v mape sa zobrazí jeho kód a názov (Obr. 33). Zo 17 regiónov je až 11 v Turecku, 4 sú v Rumunsku a 1 región je v Grécku a v Srbsku.

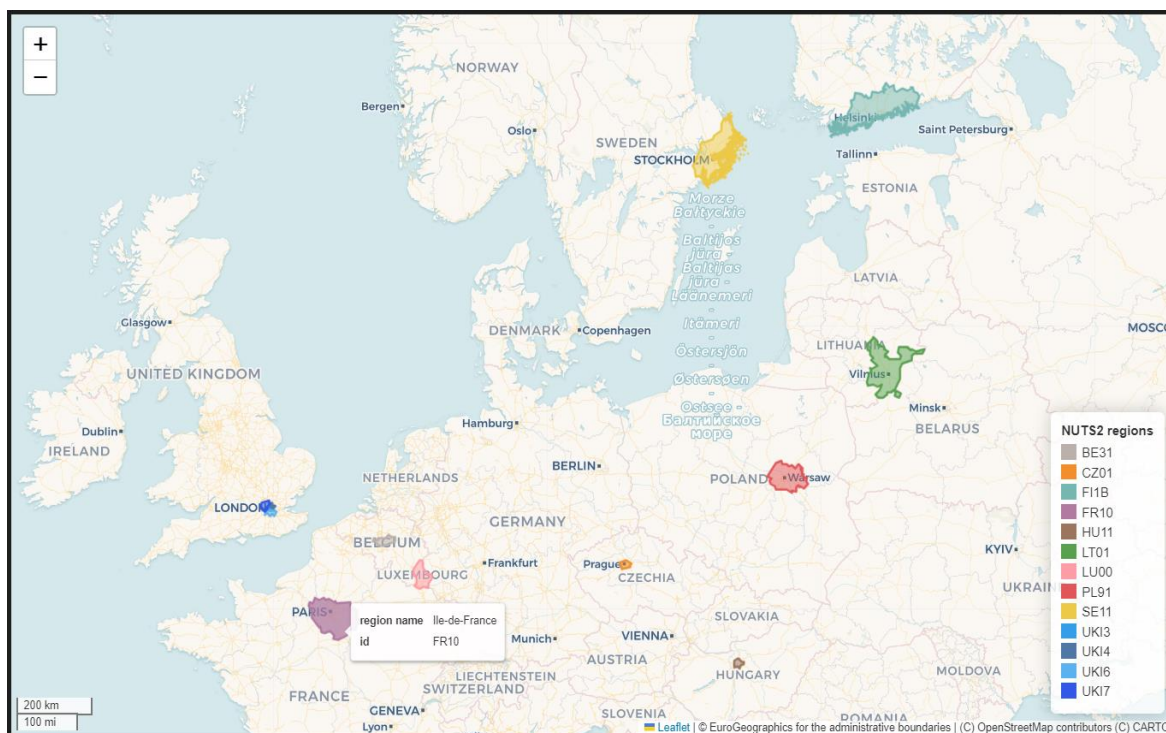


Obr. 33 Interaktívna mapa 5 regiónov s najnižšími HRST podielmi (autor: A. Tóth).

Celý postup je zopakovaný pre 5 regiónov s najvyššími hodnotami HRST. Funkcie sú už definované, teraz sa iba zavolajú s nastaveniami parametrov tak, aby bolo vybraných 5 najlepších regiónov v každom roku. Počet jedinečných regiónov, ktoré sa striedajú v najlepšej päťke, je 13, avšak počet krajín, ktorých sú súčasťou, je 10. Štyri regióny sú časti Londýna, ktoré od roku 2020 už nemajú dáta, a ostatné sú regióny hlavných miest severnej, západnej a strednej Európy. Graf je na Obr. 34 a interaktívna mapa na Obr. 35.



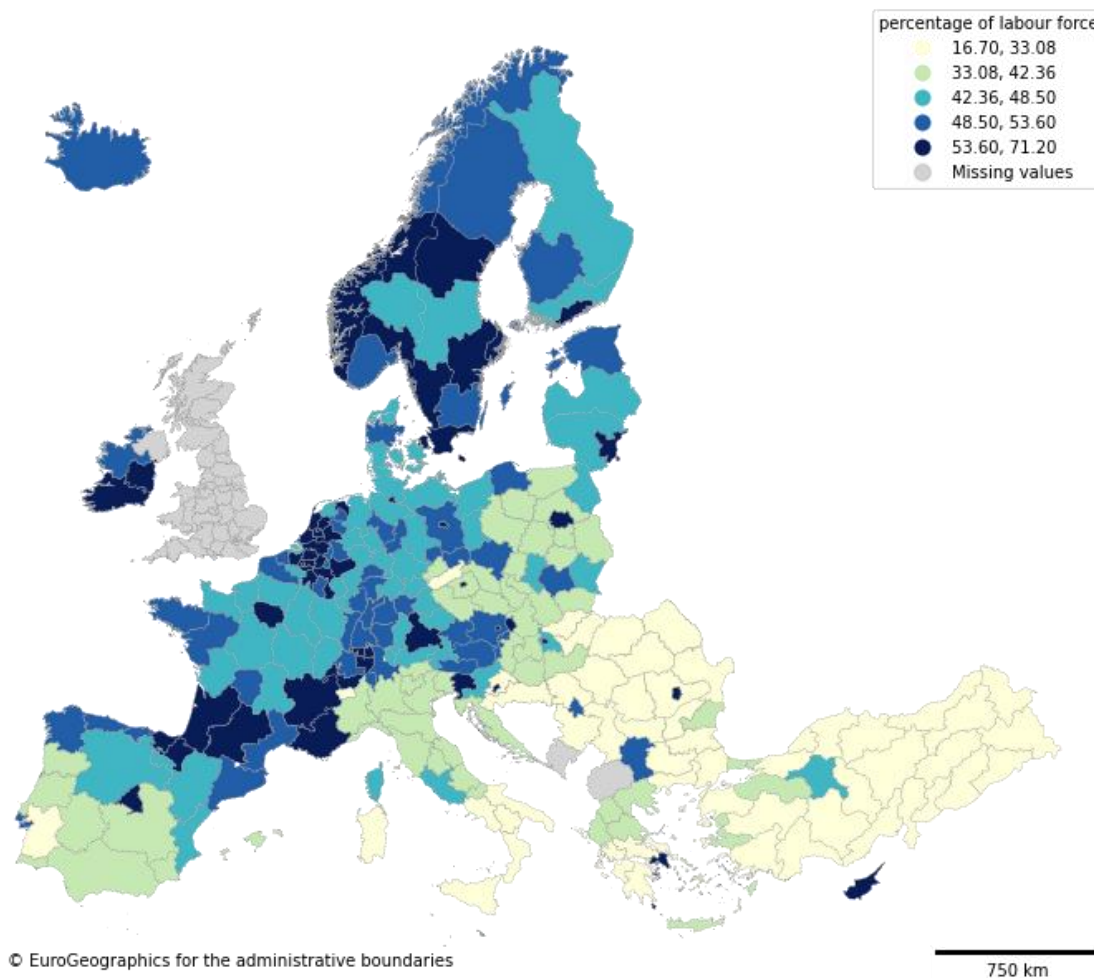
Obr. 34 Stĺpcový graf zobrazujúci 5 regiónov s najvyššími HRST podielmi (autor: A. Tóth).



Obr. 35 Interaktívna mapa 5 regiónov s najvyššími HRST podielmi (autor: A. Tóth).

V tomto okamihu má užívateľ prehľad o priemernom vývoji podielov HRST, o približnom rozdelení hodnôt a o regiónoch s najnižšími a najvyššími podielmi HRST počas sledovaného obdobia. Chýba však aktuálny prehľad všetkých regiónov. Na to je vhodný pseudokartogram, ktorý zobrazí hodnoty HRST v roku 2022. Pseudokartogram je možné vytvoriť funkciou `plot` s nastavením farebnej stupnice. Z mapy (Obr. 36) je možné vypožorovať určité rozdelenie Európy na dve časti. Severozápadná Európa má vo všeobecnosti vyššie podiely HRST ako juhovýchodná. Výnimkami sú hlavné mestá, ktoré majú vysoké hodnoty v celej Európe.

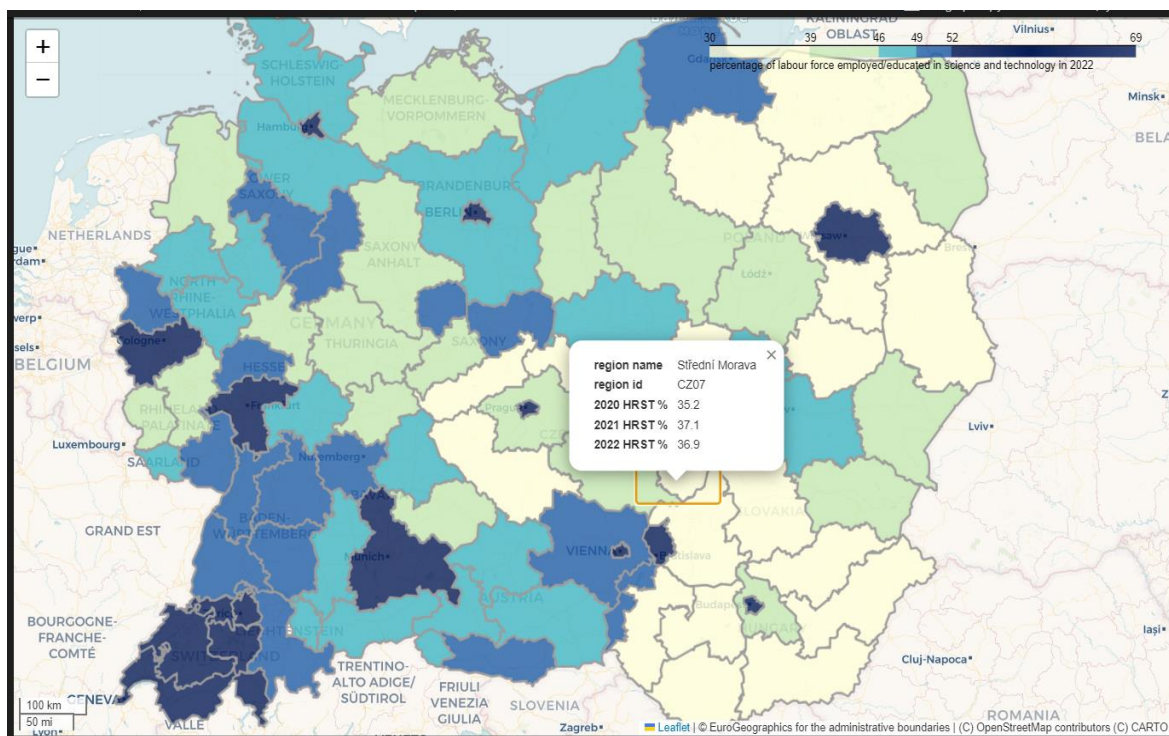
Human resources in science and technology by NUTS2 regions in 2022



Obr. 36 Statická mapa regiónov NUTS 2 s podielmi HRST v roku 2022 (autor: A. Tóth).

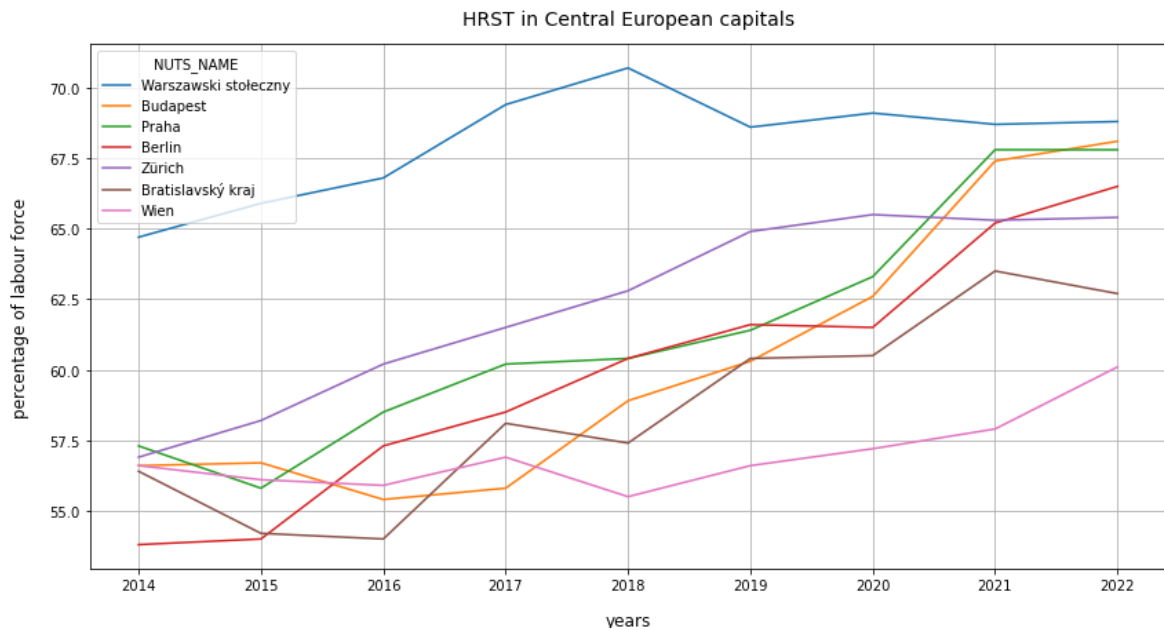
Výhodou statickej mapy vytvorenej funkciou `plot` cez knižnicu `Matplotlib` je, že vie jednoducho zobraziť geografické dáta takmer v akomkoľvek súradnicovom systéme a projekcii. Preto boli dáta HRST v celej Európe vizualizované statickou mapou. Na druhej strane interaktívna mapa vytvorená funkciou `explore` cez knižnicu `Folium` natívne podporuje iba 3 projekcie a pridanie ďalšej vyžaduje značné skúsenosti a čas. Iné možnosti a knižnice, ktoré vedú vytvoriť interaktívnu mapu, sú na tom podobne, keďže viaceré sú rovnako ako `Folium` založené na `Leaflet.js`. Knižnica `Plotly` vie generovať mapy s istou mierou interaktivity v rôznych projekciách, avšak viac sa tejto knižnici venuje notebook `Mapovacie knižnice`. Interaktívna mapa bola nakoniec vytvorená iba pre regióny strednej Európy. Stredná Európa je definovaná rôznymi spôsobmi, pre tento notebook ju však tvoria

tieto štáty: Nemecko, Švajčiarsko, Rakúsko, Maďarsko, Slovensko, Česká republika a Poľsko. Pomocou kódov týchto krajín sú vybrané ich regióny a je z nich vytvorený nový GeoDataFrame. Tento je vstupom do funkcie explore, ktorá vytvára interaktívnu mapu. Mapa (Obr. 37) je opäť pseudokartogramom hodnôt HRST za rok 2022, používa aj rovnakú farebnú stupnicu ako predchádzajúca statická mapa. Pri prejdení myšou ponad región sa zobrazí jeho názov a podiel HRST v roku 2022, po kliknutí na región sa otvorí tzv. „popup“ okno, v ktorom je názov a kód regiónu a podiely HRST v rokoch 2020, 2021 a 2022.



Obr. 37 Interaktívna mapa stredoeurópskych regiónov NUTS 2 s podielmi HRST v roku 2022 (autor: A. Tóth).

Poslednou vizualizáciou je líniový graf zobrazujúci vývoj hodnôt HRST v najväčších mestách stredoeurópskych štátov v rokoch 2014-2022. Najväčším mestom v každom štáte je hlavné mesto okrem Švajčiarska, kde je najväčším mestom Zürich, ale hlavným Bern. Ako už bolo uvedené vyššie, hlavné a najväčšie mestá v krajinách sú výnimkami v podieli HRST. Hlavným dôvodom je to, že univerzity, inštitúcie a firmy, ktoré vzdelávajú a zamestnávajú ľudí vo vedeckom a technologickom sektore, sú koncentrované do centier krajín. Týmito centrami sú najčastejšie hlavné alebo najväčšie mestá, ktoré vďaka svojej veľkej populácii tvoria samostatné regióny NUTS 2. V prípade Varšavy a Bratislavy sú do ich regiónu NUTS 2 zahrnuté aj ich aglomerácie s okolím. Z regiónov týchto najväčších miest s ich hodnotami HRST je zostavený nový DataFrame, ktorý je vstupom pre líniový graf (Obr. 38). Graf začína až od roku 2014 z dôvodu chýbajúcich dát pre Budapešť a Varšavu v rokoch 2011-2013.



Obr. 38 Líniový graf zobrazujúci vývoj podielov HRST v najväčších mestách stredoeurópskych štátov (autor: A. Tóth).

4.5 Rozloha ľadovca

Umiestnenie a názov súboru: /4 **glacier_area/glacier_area.ipynb**

Vstupné dáta: satelitné snímky Sentinel-2

Konkrétny prípad: ľadovec Aletschgletscher vo švajčiarskych Alpách

Potrebný komerčný software: nie

Počet markdown buniek: 34

Počet buniek kódu: 17

Počet riadkov kódu: 142

Úvod

Notebook vychádza zo semestrálnej práce s názvom „GeoPython: výpočet rozlohy ľadovca“ (Potočiar a Tóth 2023). Práca bola prezentovaná v máji 2023 v rámci predmetu KGI/MEDPZ (Pokročilé metódy DPZ). Autori sa v tejto práci zaoberali možnosťami GeoPythonu, čiže Python knižnicami, ktoré poskytujú GIS funkcionality bez priameho použitia GISu. Konkrétnym príkladom využitia GeoPythonu bol výpočet rozlohy ľadovca Aletschgletscher pomocou NDSI (Normalized Difference Snow Index). Tento notebook nadväzuje na Python skript z tejto práce a vylepšuje ho, dokumentuje jednotlivé kroky, ktoré vedú k samotnému výpočtu rozlohy a na záver vizualizuje výsledky. Na začiatku notebooku sa nachádza návod na stiahnutie družicových dát z misie Sentinel-2.

Rovnako ako predchádzajúci Eurostat notebook, ani tento nevyžaduje ArcGIS funkcionality ani žiaden iný komerčný software alebo licenciu. Je možné ho spúšťať a používať v akomkoľvek programe, ktorý vie pracovať s Jupyter Notebookmi (ArcGIS Pro, VS Code, JupyterLab a pod.). Python prostredie, ktoré bude fungovať ako kernel pre tento notebook, musí obsahovať týchto 7 knižníc: Rasterio (Mapbox 2018), Shapely (Gillies a

Shapely contributors 2024), NumPy (NumPy Developers 2022), OSMnx, Matplotlib, re (Python Software Foundation 2024b) a os (Python Software Foundation 2024a).

Satelitné snímky s misie Sentinel-2 potrebné pre tento notebook sa nachádzajú v zložke „data“ v adresári, kde je uložený aj tento notebook. Ak si chce užívateľ stiahnuť dáta samostatne alebo chce snímky z iného obdobia, môže tak urobiť po vytvorení si účtu v [Copernicus Data Space Ecosystem](#), ktorý je zadarmo.

NDSI a jeho využitie pre odhad rozlohy ľadovca

NDSI sa počíta ako rozdiel zeleného pásma (zelená časť viditeľného spektra) a krátkovlnného infračerveného pásma vydelení ich súčtom. Hodnoty NDSI tak môžu byť v rozsahu od -1 do 1. Pre Sentinel-2 je zelené pásmo pomenované ako Band₃ a krátkovlnné infračervené pásmo je označené ako Band₁₁ (Obr. 39).

$$\begin{aligned} NDSI_{L8} &= \frac{Band_3 - Band_6}{Band_3 + Band_6} && \text{Landsat 8 OLI} \\ NDSI_{L5} &= \frac{Band_2 - Band_5}{Band_2 + Band_5} && \text{Landsat 5 TM} \\ NDSI_{S2} &= \frac{Band_3 - Band_{11}}{Band_3 + Band_{11}} && \text{Sentinel 2 MSI} \end{aligned}$$

Obr. 39 Vzorce na výpočet NDSI pre rôzne zdroje družicových dát (autor: D. Sykas, zdroj: <https://www.geo.university/pages/blog?p=spectral-indices-with-multispectral-satellite-data>).

Sneh má najvyššiu odrazivosť elektromagnetického žiarenia vo viditeľnej časti spektra, ktorá je v NDSI reprezentovaná zeleným pásmom. Na druhej strane, sneh absorbuje väčšinu žiarenia v krátkovlnnom infračervenom pásme. To je dôvod, prečo práve tieto 2 pásma vstupujú do výpočtu – pomocou tohto indexu je jednoduché odlišiť sneh od ostatných povrchov.

V tomto momente je na mieste otázka: Prečo sa používa index na detekciu snehu, keď úlohou je detegovať ľadovec? NDSI môže byť použitý aj na detekciu ľadu a ľadovcov, ako píše D. Sykas (2020) v článku Spectral Indices with multispectral satellite data: „NDSI sa bežne používa pri mapovaní snehovej a ľadovej pokrývky a preto sa môže sa použiť aj pri monitorovaní ľadovcov“. Existuje aj index špeciálne vyvinutý na mapovanie ľadovcov: Normalized Difference Glacier Index (NDGI), ale pri tomto indexe je náročné nájsť jasné prahové hodnoty, ktoré oddelia ľadovce od ostatných povrchov. Preto bol nakoniec zvolený NDSI ako metóda na klasifikáciu ľadovca na satelitnej snímke.

Poslednou otázkou je, ako určiť prahovú hodnotu NDSI pre sneh/ľad. Odpoveď sa nachádza na webstránke [Level-2A Algorithm Overview](#): „Pixely s hodnotami NDSI presahujúcimi prahovú hodnotu NDSI T2 = 0,42, sa považujú za pixely snehu (pravdepodobnosť, že na pixeli je sneh, sa rovná 1,0)“ (ESA no date).

Metóda odhadu plochy ľadovca prezentovaná v tomto notebooku je veľmi jednoduchá a nejde o presný výpočet, ale skôr o odhad. Na dosiahnutie presnejších výsledkov je potrebné použiť lepšiu a detailnejšiu metodiku, než je len výpočet NDSI.

Stiahnutie snímok Sentinel-2

Pôvodným zámerom bolo využiť submodul SentinelAPI z knižnice Sentinelsat na stiahnutie dát z družíc Sentinel-2 podobným postupom ako v článku „Satellite imagery access and analysis in Python & Jupyter Notebooks“ (Hassan 2019). Sťahovanie s touto knižnicou je však od 2.11.2023 nefunkčné, pretože odvtedy sú Sentinel dáta dostupné na novom úložisku [Copernicus Data Space Ecosystem](#). S týmto úložiskom sa rušia staré API, ktoré boli naviazané na predchádzajúce úložisko [Copernicus Open Access Hub](#) a vznikajú nové. Nakoniec boli snímky zo Sentinel-2 stiahnuté manuálne cez [Copernicus Browser](#) a celý postup od výberu územia cez nastavenie parametrov vyhľadávania snímok až po ich samotné stiahnutie je zdokumentovaný formou obrázkov (screenshots) so stručnými komentármi v prvej časti tohto notebooku.

Oblasťou záujmu je ľadovec Aletschgletscher vo švajčiarskych Alpách. Je to najväčší alpský ľadovec, čo ho robí vhodným objektom na odhad jeho rozlohy a sledovanie zmien v čase. Okolo tohto ľadovca je na mape v prostredí Copernicus Browser nakreslený obdĺžnik, ktorý nastavuje priestorový filter na vyhľadávanie Sentinel-2 snímok. Druhým filtrom je výber zdroju dát. Tým je misia Sentinel-2, družica S2A a dáta typu L2A, čo sú snímky, ktoré prešli obrazovými korekciami vrátane atmosférickej korekcie. Posledným filtrom je definovanie obdobia, z ktorého majú snímky pochádzať. V tomto prípade je potrebné získať snímky z letného obdobia (ideálne stred júla) počas posledných 6 rokov (2018-2023). Po nastavení všetkých filtrov je spustené vyhľadávanie a vo výsledkoch je hľadaný snímok, ktorý bol spravený čo najbližšie k 15.7. a na ktorom nie je Aletschgletscher zakrytý oblakmi. Vyhovujúci snímok je stiahnutý ako zazipovaný súbor, ktorý je potrebné rozbaľiť a umiestniť do zložky „data“ v adresári, kde je uložený tento notebook.

Podľa návodu v tejto prvej časti si vie užívateľ sám stiahnuť potrebné dáta pre tento notebook alebo pre jeho ďalšie projekty. Pre urýchlenie používania tohto notebooku sú však spolu s ním šírené stiahnuté Sentinel-2 snímky v zložke „data“, konkrétne iba pásma B03 (zelená časť viditeľného spektra) a B11 (krátkovlnná časť infračerveného spektra) s 20 m priestorovým rozlíšením. Tieto pásma sú potrebné na výpočet NDSI.

Príprava dát

Prvým krokom je importovanie potrebných knižníc Rasterio, OSMnx, Shapely, re, os, NumPy a Matplotlib. Nasleduje uloženie snímok do 2 zoznamov podľa ich spektrálneho pásma. Metódou [os.walk](#) je prejdená zložka „data“ a snímky pásma B03 sú uložené do zoznamu bands_B03 a snímky pásma B11 sú uložené do zoznamu bands_B11. Do zoznamu unique_dates sú uložené dátumy vyhotovenia snímok. Tieto dátumy sú získané pomocou regulárneho výrazu, pomocou ktorého funkcia [re.search](#) hľadá dátum v názve snímku. S výnimkou roku 2018, v ktorom je Sentinel-2 snímka vyhotovená 1.7., sú za všetky roky stiahnuté snímky vyhotovené v polovici júla (rozsah dátumov je 13.-18.7.)

Na získanie polygónu ohraničujúceho ľadovec je aplikovaný rovnaký postup ako pri získavaní polygónu ohraničujúceho mesto Liverpool v notebooku Hustota dopravnej siete. Funkciou geocode to gdf z knižnice OSMnx je polygón reprezentujúci Aletschgletscher v OSM uložený ako GeoDataFrame. Aby však tento polygón správne orezal snímky zo Sentinel-2, je potrebné zjednotiť ich súradnicové systémy. Keďže je jednoduchšie a rýchlejšie preprojektovať jednoduchý polygón ako 12 rastrových súborov, súradnicový systém snímok je určený ako cieľový. Ten sa zistí pomocou otvorenia jedného snímku funkciou [rasterio.open](#) a vypísaním jeho metadát. Cieľový súradnicový systém je WGS 84/UTM zone 32N (epsg: 32632). Polygón ľadovca je teda preprojektovaný do tohto

súradnicového systému funkciou to crs. Nakoniec je z neho vytvorený tzv. [shapely polygón](#), pretože GeoDataFrame polygónom nie je možné orezávať rastrové dáta.

Pre orezanie snímok je definovaná funkcia clip_raster, ktorá má jeden vstupný parameter bands a tým je zoznam snímok. Funkcia prejde všetky snímky v zozname, oreže ich funkciou [mask](#) podľa polygónu ľadovca, aktualizuje ich metadátovej záznam a vráti dva zoznamy. Prvý zoznam obsahuje samotné orezané snímky a druhý obsahuje metadáta k orezaným snímkam. Funkcia je zavolaná 2-krát, raz pre zoznam bands_B03 a druhýkrát pre zoznam bands_B11.

Výpočet NDSI a odhad rozlohy ľadovca

Prípravou na uloženie rastrov s vypočítaným NDSI je vytvorenie zložky „ndsi“ v adresári, kde sa nachádza notebook aj zložka „data“. Na vytvorenie tejto zložky je použitá funkcia os.mkdir.

Nasleduje for cyklus, ktorý určí rozlohu ľadovca v každom sledovanom roku. Najprv je vypočítaný index NDSI a uložený do rastrovej premennej. Pixely tohto nového rastra sú reklasifikované na hodnoty 1 (ľadovec) a 0 (ostatné povrchy). Pixel je klasifikovaný ako ľadovec v prípade, že jeho hodnota NDSI je vyššia ako 0,42, čo je prahová hodnota medzi ľadom a ostatnými povrchmi. Táto prahová hodnota je špecifikovaná v dokumentácii „Level-2A Algorithm Overview“ (ESA no date). Ďalším krokom je spočítanie pixelov, ktoré po reklasifikácii majú hodnotu 1. Rozloha ľadovca je vypočítaná ako počet týchto pixelov krát veľkosť pixela vydelená jedným miliónom, aby bol výsledok v km². Keďže používané pásma majú 20 m priestorové rozlíšenie, veľkosť jedného pixela je 400 m² (20 × 20 = 400). Vypočítaná rozloha je uložená do zoznamu glacier_area, v ktorom budú po skončení cyklu uložené rozlohy ľadovca pre všetkých 6 sledovaných rokov. Na záver iterácie je vypísaná práve vypočítaná rozloha spolu s dátumom, pre ktorý je táto rozloha platná a reklasifikovaný NDSI raster je zapísaný vo formáte tif do zložky „ndsi“. Kód tohto for cyklu je na Obr. 40.

```
# define an empty list glacier area - you will store the glacier area for every year in this list
glacier_area = []

# let i take values from 0 to half of the source list length
for i in range(len(clipped_B11_image)):
    ndsi = (clipped_B03_image[i] - clipped_B11_image[i]) / (clipped_B03_image[i] + clipped_B11_image[i])

    # reclassification: pixels with NDSI greater than 0.42 are assigned value 1 = glacier pixels,
    # the rest is assigned value 0 = no glacier pixels
    ndsi[(ndsi > 0.42)] = 1.0
    ndsi[(ndsi <= 0.42)] = 0.0

    # glacier area calculation: count the number of pixels classified as glacier
    glacier_pixels_count = np.sum(ndsi == 1.0)
    # this number is multiplied by the pixel size, in this case 400 m2, because the image bands have
    # 20 m spatial resolution (20x20=400), then it's divided by 1 million so the final glacier size is in km2
    glacier_area.append((glacier_pixels_count * 400) / 1000000)
    date = unique_dates[i]
    print(f"glacier area on {date[6:]} + '.' + date[4:6] + '.' + date[:4]: {glacier_area[i]} km2")

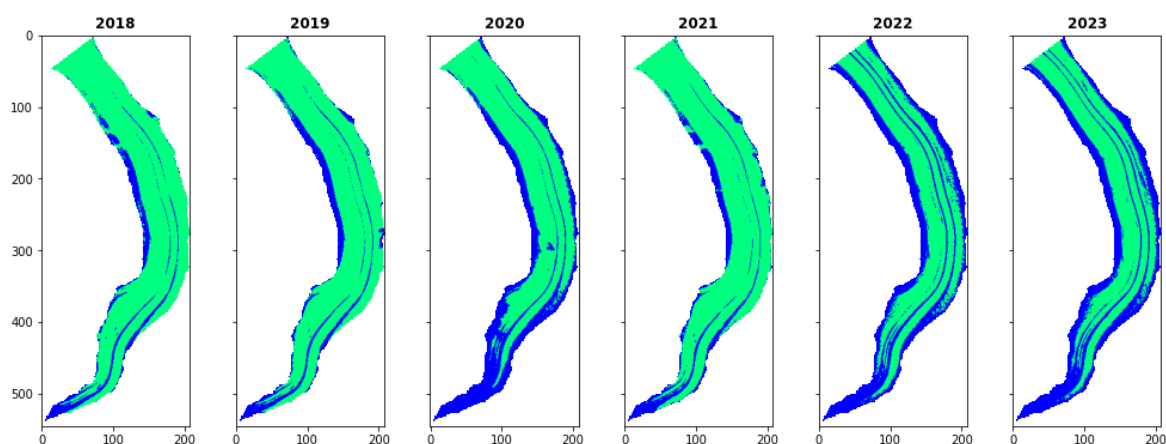
    # save calculated NDSI as a raster file named ndsi_date.tif
    with rasterio.open(cwd + f'/ndsi/ndsi_{unique_dates[i]}.tif', 'w', **clipped_B03_meta[i]) as dst:
        dst.write(ndsi)

glacier area on 01.07.2018: 12.1368 km2
glacier area on 16.07.2019: 11.6908 km2
glacier area on 13.07.2020: 9.0392 km2
glacier area on 18.07.2021: 11.5788 km2
glacier area on 13.07.2022: 8.1848 km2
glacier area on 15.07.2023: 8.0132 km2
```

Obr. 40 For cyklus, v ktorom je vypočítaná rozloha ľadovca (autor: A. Tóth).

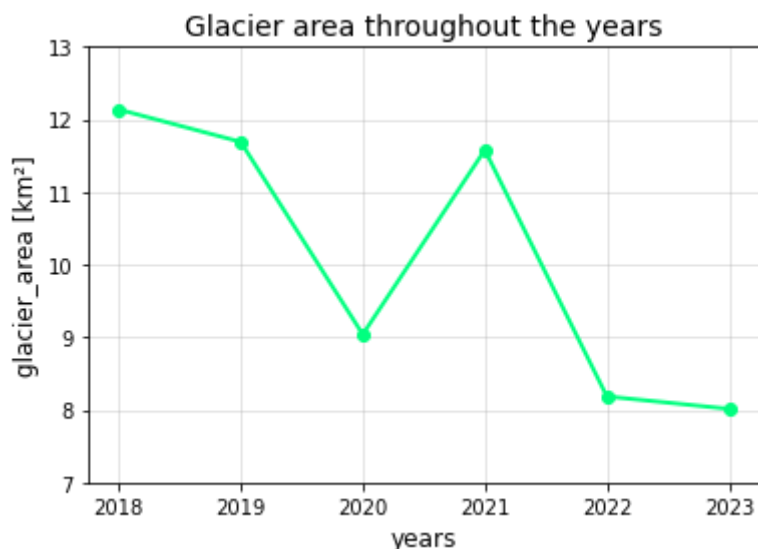
Vizualizácia zmeny rozlohy ľadovca

Prvou vizualizáciou je porovnanie reklasifikovaných NDSI rastrov zo všetkých 6 rokov, na ktorých je zelenou farbou označený ľadovec a modrou ostatné povrchy. Pri pohľade na tieto rastre je vidieť, že z ľadovca Aletschgletscher ubúda hlavne na okrajoch a v spodnej časti. Vizualizácia (Obr. 41) bola vytvorená pomocou submodulu Pyplot knižnice Matplotlib, ktorý umožňuje zobrazíť viacero grafov vedľa seba. Funkciou `rasterio.open` sú postupne otvárané NDSI rastre uložené vo formáte tif v zložke „ndsi“ a príkazom `plot.show` sú zafarbené a zobrazené v grafoch vedľa seba.



Obr. 41 Ľadovec Aletschgletscher (zelená farba) v júli v rokoch 2018-2023 (autor: A. Tóth).

Druhou a poslednou vizualizáciou je líniový graf (Obr. 42), ktorý sleduje vývoj rozlohy ľadovca počas sledovaného obdobia. Aj tento graf je vytvorený submodulom Pyplot knižnice Matplotlib. Ľadovec sa postupne stále zmenšoval, výnimkou je rok 2021, kedy bola jeho rozloha väčšia ako v roku 2020. Celková plocha, o ktorú sa ľadovec Aletschgletscher zmenšil medzi rokmi 2018 a 2023, je 4,1236 km².



Obr. 42 Líniový graf vývoja rozlohy ľadovca v júli v rokoch 2018-2023 (autor: A. Tóth).

4.6 Mapovacie knižnice

Umiestnenie a názov súboru: `/5_mapping_libraries/mapping_libraries.ipynb`

Vstupné dáta: polygónová vrstva európskych štátov stiahnutá z GISCO

Konkrétny prípad: Python knižnice Cartopy, Folium, Plotly, ipyleaflet

Potrebný komerčný software: nie

Počet markdown buniek: 41

Počet buniek kódu: 28

Počet riadkov kódu: 248

Úvod

Notebook Mapovacie knižnice je inšpirovaný článkom [6 python libraries to make beautiful maps](#) (Rozanov 2024). Autor článku v ňom stručne predstavuje knižnice Cartopy, Folium, Plotly, ipyleaflet, geemap a ridgemap a pripája ukážky kódu a máp, ktoré je možné pomocou týchto knižníc vytvoriť. Tento notebook preberá myšlienku článku a hoci sa venuje iba 4 knižniciam, predstavuje ich o niečo podrobnejšie. Každá knižnica najprv vizualizuje na mape polygónovú vrstvu štátov Európy, čím sa knižnice navzájom porovnávajú na rovnakej vrstve. Tiež bude sledovaná schopnosť zobrazit štáty na európskom svetadieli vo vhodnej projekcii. Nasledovať bude ukážka silnej stránky danej knižnice, niečoho, čo ju robí špecifickou alebo na čo je vhodná.

Podobnosť s predchádzajúcimi dvoma notebookmi je v tom, že ani v tomto prípade nie je potrebný žiadny komerčný software či účet. Je na preferencii užívateľa, aký program zvolí na prácu s notebookom a aké Python prostredie nastaví ako kernel. Jedinou podmienkou na úspešné spúšťanie kódu notebooku je nainštalovanie týchto knižníc: Cartopy (Cartopy contributors 2024), Folium, Plotly (Plotly 2024), ipyleaflet (Jupyter Development Team no date), Requests, os, GeoPandas, Matplotlib, matplotlib-scalebar, NumPy, Xarray (Xarray Developers 2024), ipywidgets (Project Jupyter 2023).

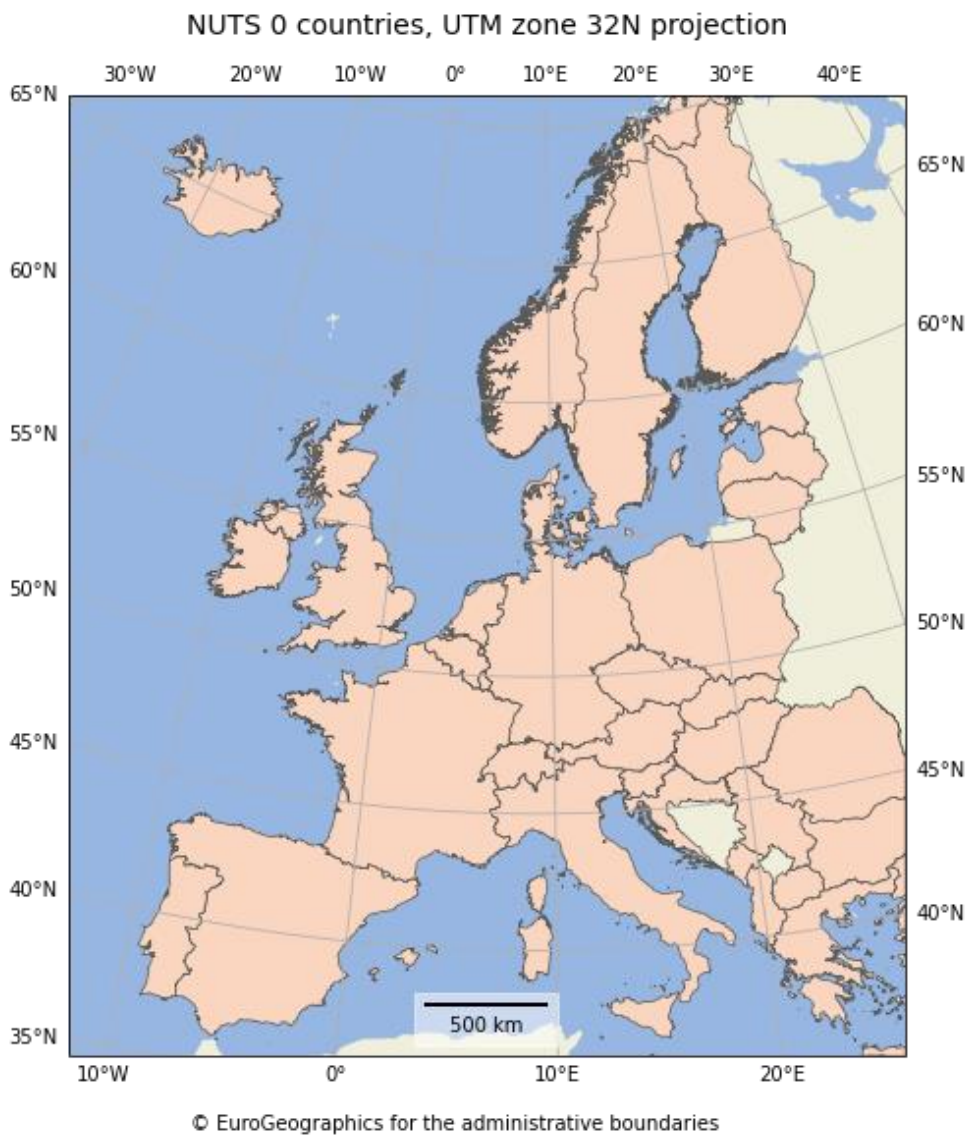
Prvým krokom je stiahnutie polygónovej vrstvy európskych štátov, ktorá bude vizualizovaná každou knižnicou. Zdrojom je GISCO a vybranou vrstvou sú regióny NUTS 0, čo sú vlastne samotné štáty, keďže level 0 je úplne prvé NUTS delenie a vždy začína na úrovni celých štátov. Vybraná vrstva je uložená vo formáte GeoJSON, je platná pre rok 2021, je uložená v súradnicovom systéme WGS 1984 (epsg: 4326) a je vytvorená v mierke 1:1 000 000/má mierkové rozlíšenie 1 milión. Stiahnutá je pomocou knižníc Requests a os rovnakým spôsobom ako bola stiahnutá vrstva regiónov NUTS 2 v notebooku Eurostat. GeoJSON súbor je uložený do adresára, kde je uložený aj tento notebook.

Následne je táto vrstva európskych štátov načítaná do štruktúry GeoDataFrame a funkcionalitou knižnice GeoPandas (funkcie [explode](#) a [dissolve](#) a súradnicový index cx) sú odstránené zámorské teritória. Nakoniec je upravená vrstva zobrazená na jednoduchej statickej mape vytvorenej funkciou plot a prvých 5 štátov je vypísaných vo forme tabuľky.

Cartopy

Cartopy je knižnicou, ktorá vie vytvárať statické mapy vo viac ako 30 kartografických zobrazeniach a transformovať medzi nimi vektorové a rastrové dáta. „Cartopy využíva knižnice PROJ, NumPy a Shapely a zahŕňa rozhranie postavené na knižnici Matplotlib pre tvorbu máp publikačnej kvality“ (Cartopy contributors 2024).

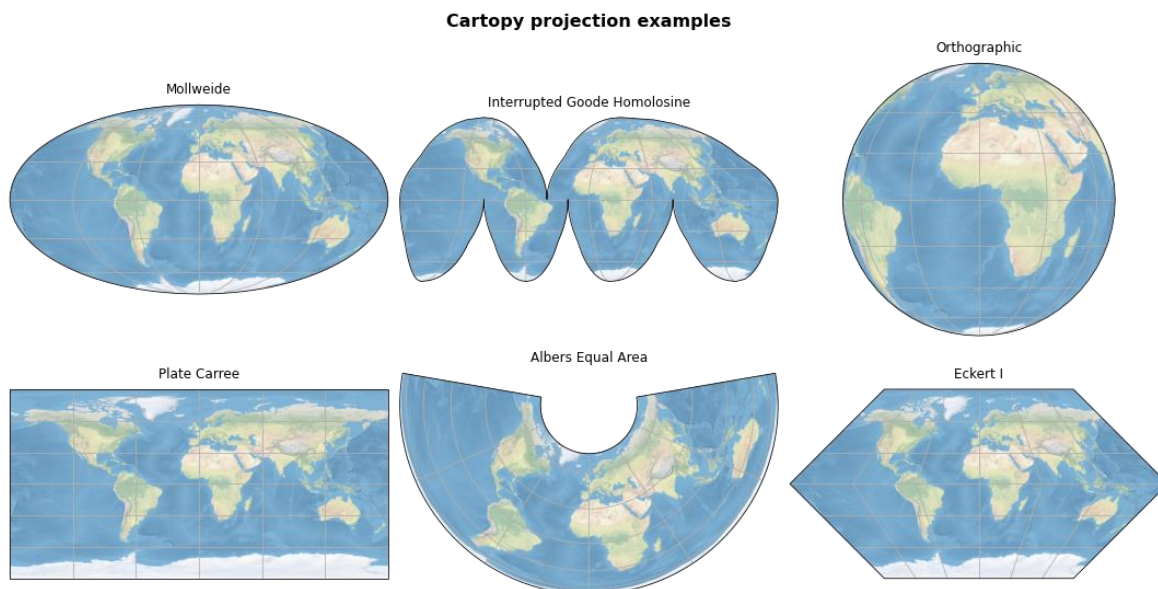
Medzi dostupnými projekciami vhodnými pre Európu je WGS 1984/UTM Zone 32N (epsg: 32632). GeoDataFrame s európskymi štátmi je preprojektovaný do tohto súradnicového systému a následne pridaný do mapy funkciou [add_geometries](#). Táto funkcia poskytuje parametre na nastavenie štýlu ako farba výplne, farba a hrúbka obrysovej línie a pod. Pod vrstvou európskych štátov je podkladová vrstva LAND z Cartopy feature interface. Táto podkladová vrstva reprezentuje všetku súš na Zemi. Grafická mierka je do mapy pridaná pomocou knižnice matplotlib-scalebar. Tiež je do mapy pridaná súradnicová sieť s popismi geografickej šírky po 5° a s popismi geografickej dĺžky po 10°. Výsledná mapa je na Obr. 43.



Obr. 43 Cartopy statická mapa európskych štátov (autor: A. Tóth).

Ako už bolo spomenuté, výhodou Cartopy je množstvo kartografických zobrazení. V nasledujúcom príklade je ukázaných 6 zaujímavých a užitočných zobrazení, ktoré Cartopy ponúka. Celý zoznam projekcií sa nachádza na [webovej stránke](#) dokumentácie knižnice. Vybrané zobrazenia sú uložené do zoznamu, ktorý vstupuje do for cyklu vykresľujúceho mapu pre každé zobrazenie. Každá mapa obsahuje jednoduchú fyzikogeografickú podkladovú vrstvu (funkcia [stock_img](#)) a súradnicovú sieť (funkcia [gridlines](#)). Pomocou

submodulu Pyplot knižnice Matplotlib sú mapy (Obr. 44) usporiadané do 2 riadkov a 3 stĺpcov.



Obr. 44 Ukážka 6 kartografických zobrazení dostupných v knižnici Cartopy (autor: A. Tóth).

Ďalšou silnou stránkou tejto knižnice sú jej podkladové vrstvy z rozhrania [Cartopy feature interface](#). Sú to líniové vrstvy hraníc, pobrežnej čiary a riek a polygónové vrstvy krajín, jazier, súše a oceánu. Využitím týchto zabudovaných vrstiev získavajú mapy priestorový kontext. Mapa (Obr. 45) s ukážkou podkladových vrstiev je v Robinsonovom zobrazení. Kód bol inšpirovaný [príkladom](#) z galérie z webovej stránky dokumentácie knižnice Cartopy.



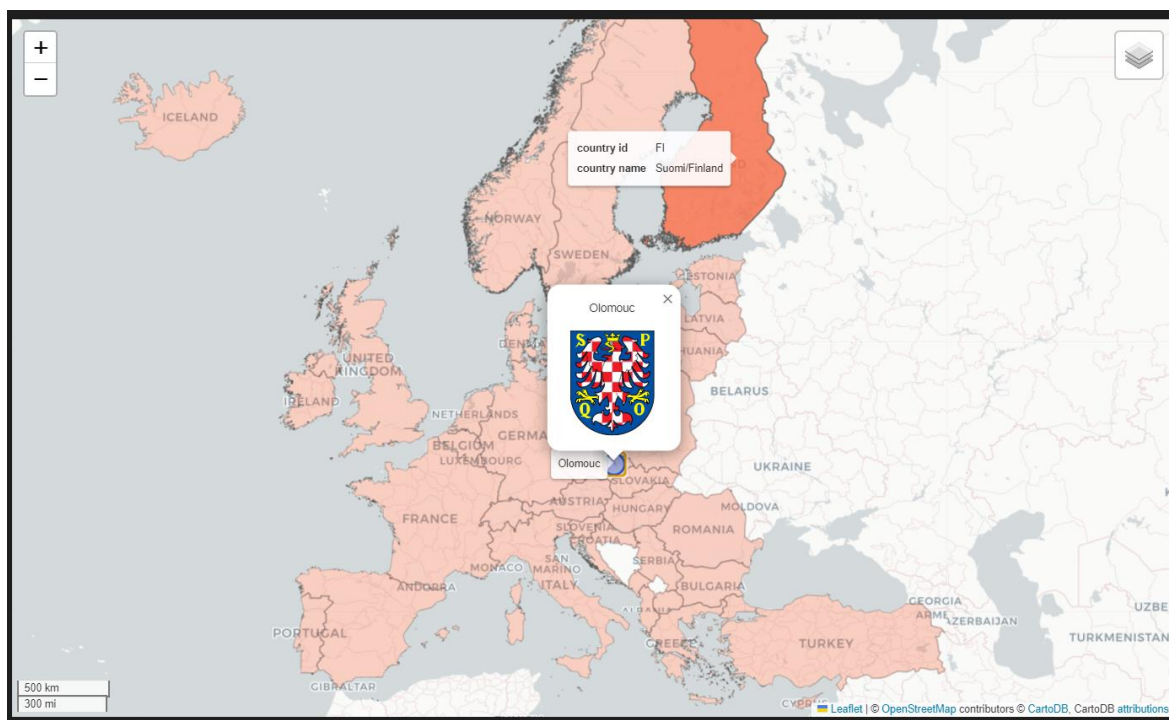
Obr. 45 Podkladové vrstvy z Cartopy feature interface (autor: A. Tóth).

Folium

Folium je jedna z najpopulárnejších mapovacích knižníc v Pythone. Vytvára interaktívne mapy postavené na JavaScriptovej mapovacej knižnici Leaflet.js. Mottom knižnice Folium je „spracuj si dáta v Pythone a potom ich vizualizuj v Leaflet mape prostredníctvom Folia“ (Story 2024). Jednou z možností je tiež uložiť si vytvorenú mapu ako HTML súbor a pokračovať vo vývoji mapy priamo pomocou knižnice Leaflet úpravou JavaScriptového a HTML kódu.

Prvým krokom je vytvorenie mapy vizualizujúcej vrstvu európskych štátov. Nevýhodou oproti knižnici Cartopy je obmedzený počet podporovaných projekcií. Konkrétne to sú tieto 3: najčastejšie podporované zobrazenie vo webových mapách WGS 84/Pseudo-Mercator (epsg: 3857), WGS 1984 (epsg: 4326) a WGS 84/World Mercator (epsg: 3395). Ani jedna však nie je vhodná pre celé územie Európy, pretože vo vyšších geografických šírkach dochádza k značnému plošnému skresleniu a štáty Nórsko, Švédsko a Fínsko sa javia oveľa väčšie, ako v skutočnosti sú. Nakoniec je pre tento prípad vybraná projekcia WGS 84/Pseudo-Mercator.

Podkladovou mapou je dlaždicová vrstva CartoDB Positron, čo je jednoduchá svetlá podkladová vrstva s riekami, hranicami štátov, ich krajov a najväčších miest. GeoDataFrame, v ktorom sú uložené európske štáty, je do Folium mapy pridaný funkciou [GeoJson](#). V tejto funkcii je nastavený štýl vykreslenia polygónov a sú definované akcie, ktoré sa majú vykonať pri interakcii užívateľa s mapou. Konkrétne ide o priblíženie a vycentrovanie mapy po kliknutí na štát a zvýraznenie štátu a zobrazenie jeho ID a názvu pri prejení myšou. Do mapy je pridaná bodová kruhová značka ([Circle Marker](#)), ktorá je typická pre Leaflet mapy. V tomto prípade reprezentuje mesto Olomouc a po kliknutí na ňu sa otvorí tzv. „popup“ okno, v ktorom sa nachádza znak mesta. Ukážka výslednej mapy je na Obr. 46.

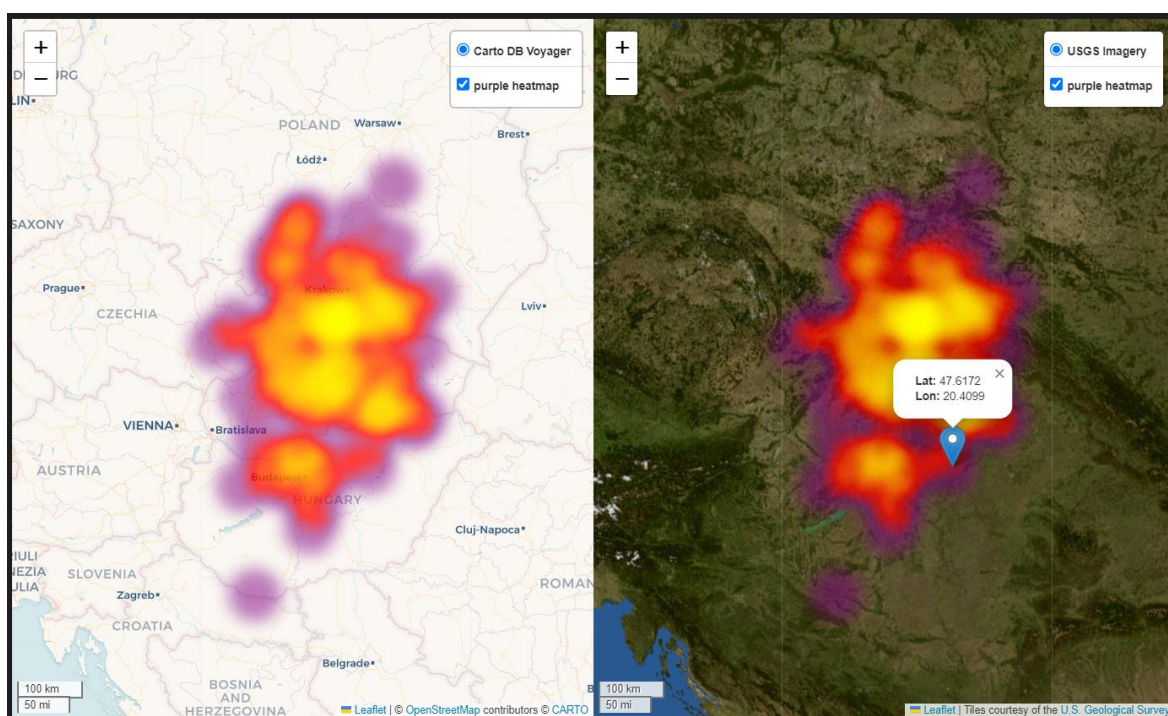


Obr. 46 Folium interaktívna mapa európskych štátov s bodovým znakom mesta Olomouc (autor: A. Tóth).

Folium obsahuje viacero zaujímavých pluginov, ich celý zoznam je dostupný na [webovej stránke](#) s dokumentáciou. V tomto notebooku sú predstavené 2 z nich, konkrétne [DualMap](#) a [HeatMap](#). Na začiatku je definované duálne mapové okno pomocou pluginu DualMap. Vlastnosťou takéhoto okna je synchronizácia interaktivity 2 vedľa seba umiestnených máp, čo znamená, že napríklad pri priblížení, oddialení, posunutí v jednej mape sa rovnako priblíži, oddialí, posunie aj druhá mapa. Mapa vľavo má ako podkladovú mapu nastavenú dlaždicovú vrstvu CartoDB Voyager, mapa vpravo má za podkladovú mapu nastavené satelitné snímky od Geologickej služby Spojených štátov (USGS).

Do oboch týchto máp je pridaná vrstva náhodne vygenerovaných súradníc na území Slovenska, ČR, Poľska a Maďarska vizualizovaná pluginom HeatMap. V tomto prípade ide len o ukážku a preto boli použité náhodne vygenerované body ako vstupné dáta. Plugin HeatMap vytvára tzv. „teplotú“ alebo „intenzitnú“ mapu, čo je vo svojej podstate raster dynamicky sa meniaci na základe úrovne priblíženia. Čím intenzívnejšie farby, tým je hustota výskytu daného javu vyššia. Plugin HeatMap umožňuje nastaviť vlastnú farebnú stupnicu, v tomto prípade boli zvolené farby fialová, červená a žltá pre intervaly 0,5, 0,7 a 1.

Na záver je do duálneho mapového okna (Obr. 47) pridaná trieda [ClickForMarker](#), ktorá vytvorí bodovú značku na mieste, na ktoré klikol užívateľ. Po kliknutí na vytvorenú bodovú značku sa zobrazí „popup“ okno, v ktorom sú geografické súradnice daného bodu.



Obr. 47 Ukážka pluginov DualMap a HeatMap a triedy ClickForMarker (autor: A. Tóth).

Plotly

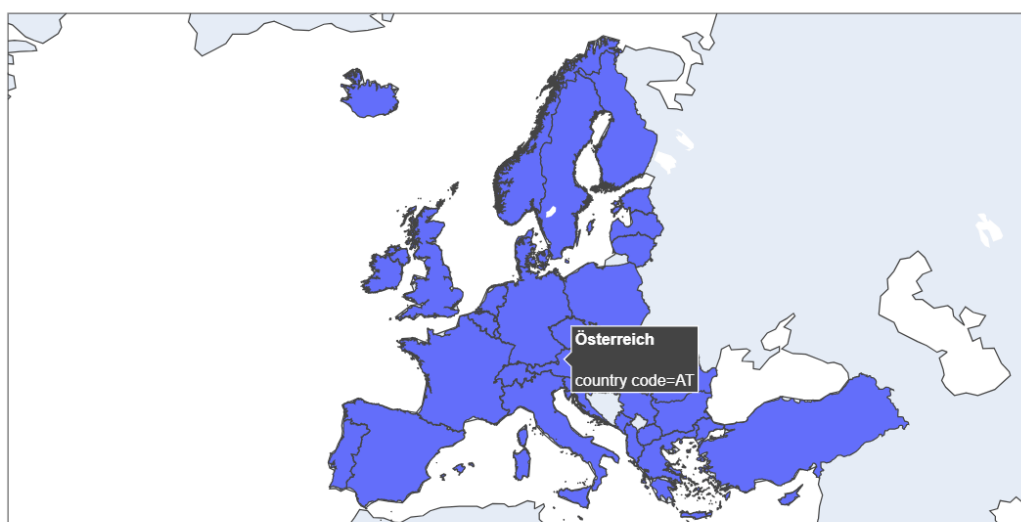
Plotly je firma, ktorá vytvára software Dash Enterprise, čo je dátová aplikačná platforma pre Python. Okrem toho združuje skupinu 8 open-source knižníc určených na vytváranie grafov a vizualizáciu dát v jazykoch Python, R, JavaScript, Julia, MATLAB. Plotly Python knižnica vie produkovať interaktívne grafy a mapy postavené na Matplotlib jadre. Ponúka

pestrú paletu 22 kartografických zobrazení. Tento notebook sa zaoberá submodulom [plotly.express](#).

V prvom kroku je funkciou [choropleth](#) vytvorená mapa európskych štátov uložených v štruktúre GeoDataFrame. Samotná funkcia ponúka množstvo parametrov, ktoré sú dobre popísané v dokumentácii. Unikátom knižnice Plotly je dynamické prispôsobovanie projekcie podľa aktuálneho územia. Užívateľ to môže otestovať na tomto príklade. Mapa je v projekcii UTM (Universal Transverse Mercator). Pri posunutí mapy na východ alebo na západ sa zobrazenie prispôsobí územi, ktoré je v strede mapy. Po prvom „načítaní“ mapy je zobrazenie pravdepodobne UTM Zone 32N, ale pri posune na východ sa „zmení“ na UTM Zone 33N atď. Podkladovou vrstvou sú pobrežné čiary a do siva zafarbené polygóny pevniny. Mapu je možné približovať, odďaľovať a posúvať, nevýhodou je však pomalá reakcia mapy na tieto podnety, čo z užívateľského hľadiska nie je príjemné. Po prejdení myšou ponad štát sa zobrazí jeho názov a kód, ako je to vidieť na Obr. 48.



Plotly: NUTS 0 countries in Transverse Mercator projection



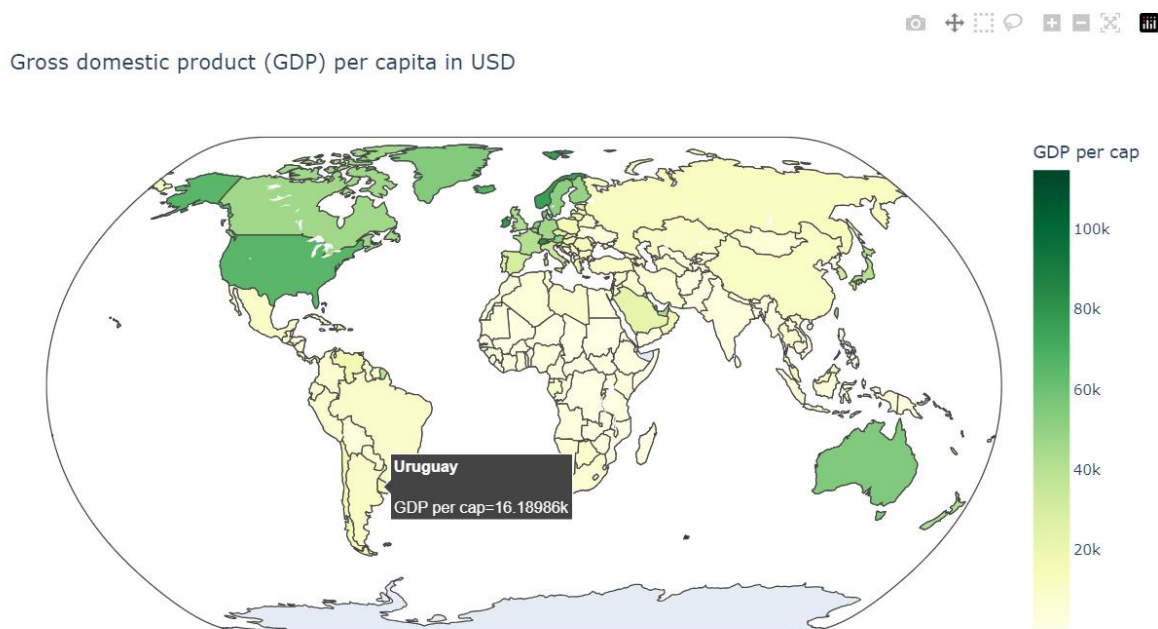
Obr. 48 Plotly interaktívna mapa európskych štátov (autor: A. Tóth).

Ako napovedá názov funkcie `choropleth`, je to funkcia určená na tvorbu kartogramov. V predchádzajúcom prípade bola použitá iba na jednoduché zobrazenie polygónovej vrstvy, v tomto ďalšom prípade ňou bude vytvorený naozajstný kartogram. Keďže však vrstva európskych štátov neobsahuje dáta, ktoré by sa dali vizualizovať kartogramom, je potrebné takéto dáta získať. Knižnica GeoPandas obsahuje submodule Datasets, v ktorom sa nachádza dataset `naturalearth_lowres`. Je to polygónová vrstva štátov sveta, ich závislých území a ďalších území na Zemi a obsahuje atribúty s počtom obyvateľom a GDP (hrubým domácim produktom). Táto vrstva je načítaná do štruktúry GeoDataFrame. Následne sú z nej odstránené územia, ktoré nemajú stálu populáciu a celé územie Antarktídy. Vo vrstve je vytvorený nový stĺpec s názvom `gdp_per_cap`, do ktorého sa vypočíta hrubý domáci produkt na obyvateľa v USD (americký dolár). Toto bude premenná, ktorá bude vizualizovaná metódou kartogramu, resp. pseudokartogramu, pretože je to veličina vzťahovaná k niečomu inému ako k ploche územia. GeoDataFrame je vypísaný vo forme tabuľky, aby si užívateľ mohol spraviť lepšiu predstavu o dátach.

Submodul `plotly.express.colors` ponúka farebné stupnice, ktoré sa dajú použiť v grafoch a mapách. Funkciou [named_colorscales](#) je možné získať názvy všetkých týchto

stupníc a vybrať si vhodnú stupnicu pre pseudokartogram. Vybraná stupnica sa volá „ylgn“ a prechádza zo svetložltých farieb do tmavozelených.

Mapa je v Natural Earth projekcii (esri: 53077) a oproti mape európskych štátov je práca s mapou (približovanie, posúvanie) rýchlejšia a plynulejšia. Po prejení myšou ponad krajinu sa zobrazí jej názov a GDP na obyvateľa v USD. Spojitá legenda je umiestnená napravo od mapového poľa. Ukážka mapy je na Obr. 49.

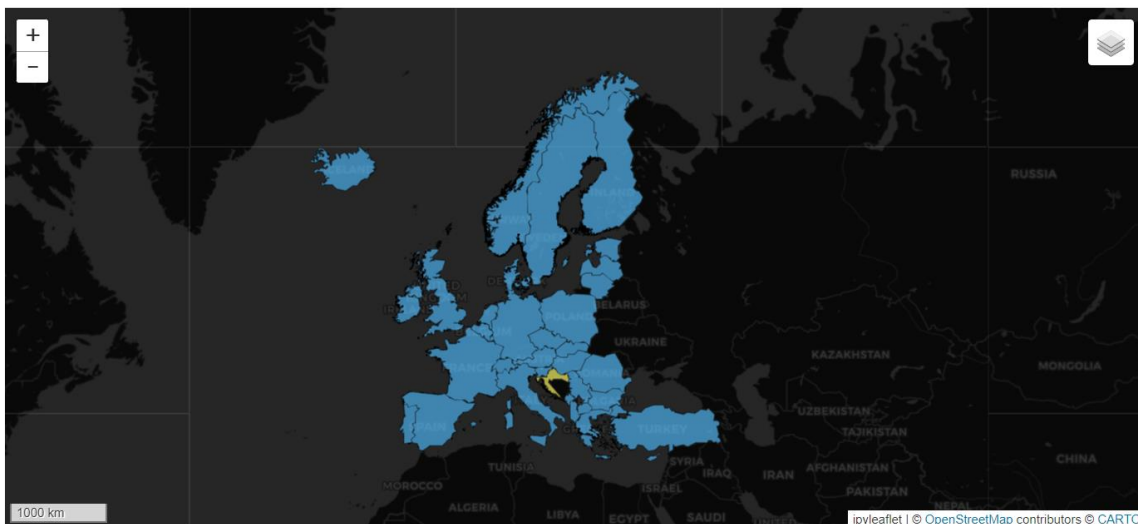


Obr. 49 Plotly pseudokartogram GDP na obyvateľa v USD (autor: A. Tóth).

ipyleaflet

Knižnica ipyleaflet je podobná Foliu v tom, že používa Leaflet.js na tvorbu interaktívnych máp. Avšak ipyleaflet patrí do skupiny Jupyter widgetov a preto môže byť táto knižnica používaná iba v Jupyter/ArcGIS Notebookoch.

Na prvej mape sú vizualizované európske štáty vo farbách EÚ: základná výplňová farba polygónov je modrá a po prejení myšou ponad ne sa zmení na žltú. Podkladovou mapou je dlaždicová vrstva CartoDB Dark Matter, čo je jednoduchá tmavá podkladová vrstva s riečnou sieťou, hranicami štátov, krajov a väčšími mestami. S kartografickými zobrazeniami je to rovnaké ako pri knižnici Folium, podporované sú iba 3 projekcie, takže pre túto mapu je zvolená základná projekcia WGS 84/Pseudo-Mercator (epsg: 3857). GeoDataFrame s európskymi štátmi je pridaný do mapy funkciou [GeoData](#), v čom je rozdiel oproti Foliu, pretože Folium používa na pridanie GeoJSON súboru aj štruktúru GeoDataFrame jednu funkciu GeoJson, zatiaľ čo ipyleaflet má 2 samostatné funkcie. Ukážka mapy je na Obr. 50.



Obr. 50 Interaktívna mapa európskych štátov vytvorená v knižnici ipyleaflet (autor: A. Tóth).

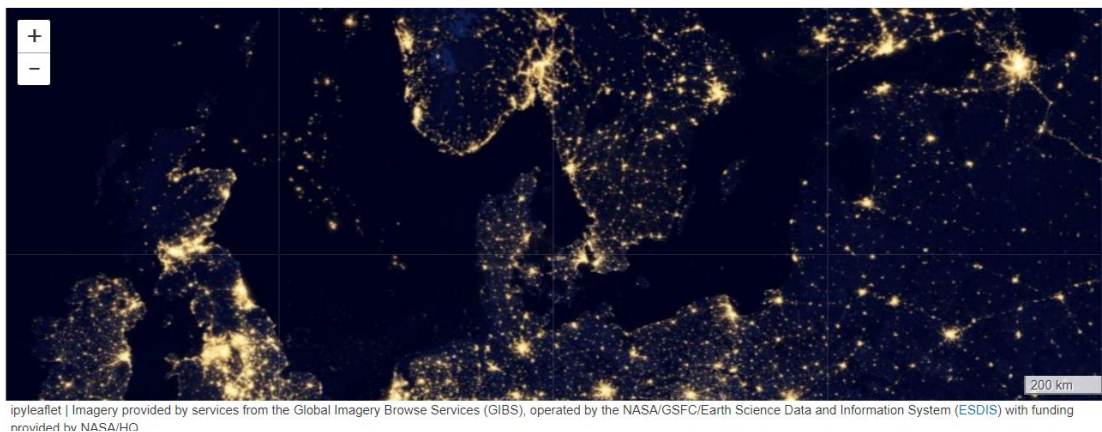
Silnou stránkou knižnice ipyleaflet je veľký výber zabudovaných podkladových máp. Zatiaľ čo Folium má zabudované iba 4 podkladové mapy, ipyleaflet ich ponúka 25. Samozrejmosťou je vloženie vlastnej dlaždicovej vrstvy alebo inej podkladovej mapy napr. od poskytovateľov [xyz services](#) alebo [leaflet providers](#). Na nasledujúcich obrázkoch (Obr. 51, Obr. 52, Obr. 53) sú vybrané 3 zaujímavé podkladové mapy, ktoré sú zabudované priamo v knižnici ipyleaflet v submodule basemaps.



Obr. 51 Podkladová mapa Esri imagery (autor: A. Tóth).



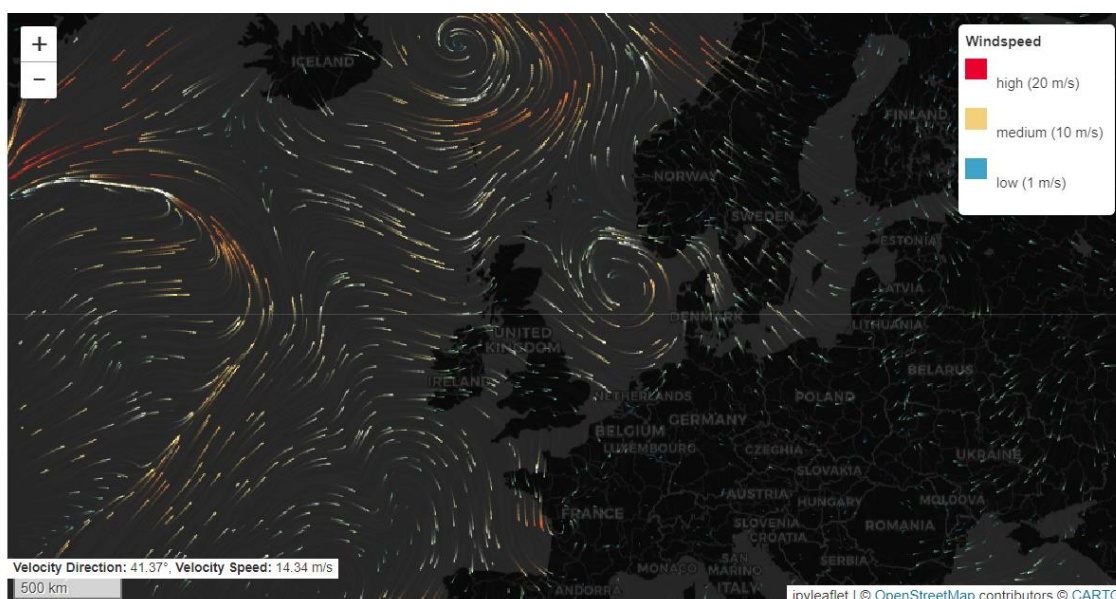
Obr. 52 Podkladová mapa Strava bicycle ride (autor: A. Tóth).



Obr. 53 Podkladová mapa NASA Earth at night 2012 (autor: A. Tóth).

Poslednou ukážkou bude dynamická vizualizácia rýchlosti a smeru vetra. Kód je prevzatý z článku „6 python libraries to make beautiful maps“ (Rozanov 2024).

Najprv je potrebné importovať knižnicu Xarray a submodul [velocity](#) z knižnice ipyleaflet. Ďalším krokom je stiahnutie dát o rýchlosti a smere vetra. Tie sa nachádzajú v súbore wind-global.nc a sú uložené na [GitHubu](#), odkiaľ sú stiahnuté podobným postupom aký bol použitý pri sťahovaní GeoJSON súborov s regiónmi NUTS 2 a 0. Na URL adresu GitHubu, kde sa súbor nachádza, sa pošle http get požiadavka pomocou knižnice Requests a odpoveď na túto požiadavku sa zapíše do adresára, kde je uložený aj tento notebook. Stiahnutý súbor wind-global.nc je prečítaný funkciou [open_dataset](#) z knižnice Xarray a stáva sa vstupom pre vrstvu velocity (rýchlosť). Táto vrstva reprezentuje vietor a vytvára dynamickú, resp. animovanú vizualizáciu, ktorá navodzuje dojem, že vietor sa naozaj hýbe a že „v mape fúka“. Rýchlosť vetra je symbolizovaná farebnou stupnicou od modrej cez žltú až po červenú, pričom červená zastupuje najväčšie rýchlosti vetra okolo 20 m/s. Smer vetra určujú animované krivky, ktoré sa „hýbu“ v smere vetra. Za podkladovú mapu je zvolená dlaždicová vrstva CartoDB Dark Matter. Pri pohybe myšou po mape sa v ľavom dolnom rohu vypisuje smer vetra ako azimutový uhol a rýchlosť vetra v m/s pre miesto, kde sa práve kurzor myši nachádza. Statické zachytenie animovanej vizualizácie je na Obr. 54.



Obr. 54 Animovaná vizualizácia smeru a rýchlosti vetra v knižnici ipyleaflet (autor: A. Tóth).

5 VÝSLEDKY

Výsledkami tejto práce sú naprogramované a napísané Jupyter Notebooky popísané v 4. kapitole. Sú dostupné na GitHub účte autora tejto diplomovej práce v repozitári [ArcGIS Notebooks thesis](#). Nasleduje popis cieľov a účelov, ktoré plnia jednotlivé notebooky. Za týmto popisom sa nachádza tabuľka s prehľadom informácií o notebookoch.

Cieľom notebooku s názvom **Základy Jupyter Notebooku** (podkapitola 4.1.1) je vysvetliť užívateľovi začínajúcemu pracovať s Jupyter Notebookmi, s čím sa chystá pracovať, na čo sa dá notebook používať a ako ho vytvoriť v prostredí VS Code. Tento notebook plní úlohu prezentácie použiteľnej vo výuke a tiež úlohu tutoriálu s radami na formátovanie textu v markdown bunkách, na prácu s premennými v bunkách kódu a s užitočnými príkazmi a funkciami v jazyku Python.

Notebook **Python cvičenia** (podkapitola 4.1.2) má za cieľ precvičiť základné programovacie znalosti z jazyka Python a praktické použitie príkazov a funkcií uvedených v predchádzajúcom notebooku. Umiestnenie riešení na koniec notebooku do zbalenej hierarchickej štruktúry necháva priestor užívateľovi vypracovať cvičenie samostatne a správne riešenie použiť iba na kontrolu. Tento notebook je možné použiť ako cvičenie vo výuke alebo ako domácu úlohu.

Cieľom notebooku **ArcPy cvičenia** (podkapitola 4.1.3) je precvičiť používanie funkcií ArcPy. Správne riešenia sú rovnako ako v predchádzajúcom notebooku umiestnené na koniec do zbalenej štruktúry. Tento notebook je špecifický tým, že jeho prvá polovica je určená na používanie vo VS Code a druhá polovica v ArcGIS Pro. Na začiatku sa nachádza návod na naklonovanie ArcGIS Python prostredia, vďaka ktorému je možné spúšťať ArcPy funkcie mimo prostredia ArcGIS Pro. Tento notebook je tak zároveň tutoriálom aj materiálom na cvičenie vo výuke.

Notebook **Kolokačná analýza** (podkapitola 4.2) má za cieľ vysvetliť kolokačnú analýzu v ArcGIS Pro. Notebook vďaka kombinácii vysvetľujúceho textu, Python kódu a okamžitých výsledkov viditeľných v mape alebo pod spustenou bunkou interaktívne vysvetľuje komplexný pojem. Užívateľ vďaka tomu získa živšiu predstavu o tom, čo kolokačná analýza je, ako jednotlivé parametre ovplyvňujú výsledky analýzy a ako sa dajú tieto výsledky interpretovať. Notebook predstavuje „data storytelling“ kolokácie mestskej zástavby so zeleňou a športoviskami v konurbácii Katowice a tiež demonštráciu mapového modulu arcpy.mp, ktorým je možné ovládať prvky a objekty v projekte ArcGIS Pro. Ako zdrojové dáta sú použité dáta Copernicus Urban Atlas pre Európu.

Cieľom notebooku **Hustota dopravnej siete** (podkapitola 4.3) je previesť užívateľa procesom výpočtu hustoty dopravnej siete na plochu a na populáciu v hexagónovej mriežke na zvolenom území. Celý proces je realizovaný pomocou Python kódu od stiahnutia dát dopravnej siete a populačného gridu, cez ich úpravu až po samotný výpočet a jednoduchú vizualizáciu. Užívateľ získa predstavu o tom, čo takýto proces zahŕňa a ako je konkrétne realizovaný v ArcGIS Pro. Zároveň tento notebook demonštruje prevod pôvodného Python skriptu do formy Jupyter Notebooku.

Notebook s názvom **Eurostat** (podkapitola 4.4) má za cieľ predstaviť knižnicu Eurostat, popisnými štatistikami charakterizovať vybraný dataset o ľudských zdrojoch vo vede a technológii a pomocou máp a grafov vytiahnuť zaujímavé informácie z dát. Pomocou uvedenej knižnice je možné jednoducho pristupovať k Eurostat datasetom na základe ich ID. Týmto spôsobom je zabezpečená aktuálnosť dát pri každom spustení notebooku. Dáta sú ukladané do štruktúr (Geo)DataFrame, takže užívateľ si osvojí ich vlastnosti a metódy. „Data storytelling“ je zameraný na vývoj podielu aktívnej populácie vo vede a technológii

v európskych regiónoch NUTS 2 medzi rokmi 2011 a 2022 a na vybrané skupiny regiónov. Čiastočným cieľom je ukážka získania, spracovania a vizualizácie priestorových dát iba s použitím open-source Python knižníc.

Cieľom notebooku s názvom **Rozloha ľadovca** (podkapitola 4.5) je predstaviť jednoduchú metódu na odhad rozlohy horského ľadovca pomocou spektrálneho indexu NDSI a demonštrovať využitie knižnice Rasterio na prácu so satelitnými snímkami. Rozloha ľadovca je odhadovaná zo Sentinel-2 snímkov stiahnutých v júli v rokoch 2018-2023. Užívateľ môže sledovať vývoj rozlohy ľadovca v tomto období na vytvorených grafoch. Čiastočným cieľom je podobne ako pri notebooku Hustota dopravnej siete konverzia zdrojového Python skriptu na Jupyter Notebook. V úvode notebooku sa nachádza návod na stiahnutie snímkov z misie Sentinel-2.

Notebook **Mapovacie knižnice** (podkapitola 4.6) má za cieľ predstaviť štyri open-source Python knižnice, ktoré vedú vytvárať mapy, ukázať silné stránky a špeciality týchto knižníc a porovnať ich pri vizualizácii jedného spoločného datasetu. Toto sú vybrané knižnice: Cartopy, Folium, Plotly a ipyleaflet. Užívateľ tak získa prehľad o možnostiach vytvárať statické i interaktívne mapy priamo v Jupyter Notebooku pomocou Python kódu. Spoločným datasetom, ktorý vizualizujú všetky knižnice, je polygónová vrstva európskych štátov.

Tabuľka 2 Prehľad informácií o vytvorených notebookoch

Názov notebooku	Počet markdown buniek	Počet buniek kódu	Počet riadkov kódu	Zdroj vstupných dát	Získanie dát	Potreba ArcGIS Pro
<i>Základy Jupyter Notebooku</i>	50	18	40	-	-	Nie
<i>Python cvičenia</i>	19	8	70	-	-	Nie
<i>ArcPy cvičenia</i>	38	11	59	ArcČR 500 verzia 3.3	Stiahnutie vopred	Áno
<i>Kolokačná analýza</i>	79	50	270	Urban Atlas LULC 2018	Registrácia a stiahnutie vopred	Áno
<i>Hustota dopravnej siete</i>	64	44	247	OSM, GEOSTAT populačný grid	Stiahnutie priamo v kóde	Áno
<i>Eurostat</i>	75	57	334	Eurostat, GISCO	Stiahnutie priamo v kóde	Nie
<i>Rozloha ľadovca</i>	34	17	142	Sentinel-2	Registrácia a stiahnutie vopred	Nie
<i>Mapovacie knižnice</i>	41	28	248	GISCO	Stiahnutie priamo v kóde	Nie

6 DISKUSIA

Limity a nedostatky

Výhodou ArcGIS Notebookov je ich integrácia s ArcGIS Online a ArcGIS API pre Python. Vizualizácia priestorových dát v notebooku môže byť realizovaná interaktívnymi mapovými oknami s ArcGIS Online mapami, ArcGIS API pre Python umožňuje publikovať vlastné vrstvy do ArcGIS Online úložiska, využívať Living Atlas ako zdroj dát a zdieľať notebooky naprieč organizáciou. Tieto možnosti sú však dostupné iba pre užívateľov, ktorých účet má na to príslušnú extenziu licencie. Táto extenzia však nie je zahrnutá v licencií univerzitného študentského ArcGIS účtu, ktorým disponuje autor tejto diplomovej práce. Preto bolo využitie ArcGIS Notebookov limitované na prostredie ArcGIS Pro, knižnicu ArcPy a na ďalšie open-source knižnice, prostredia a dátové zdroje.

Z piatich hlavných notebookov dva notebooky Kolokačná analýza a Hustota dopravných sietí demonštrujú funkčné prepojenie ArcGIS Pro a ArcPy s open-source knižnicami ako Pandas, GeoPandas, Matplotlib, Folium, OSMnx, Requests a pod. Takýchto notebookov mohlo byť viac, avšak ostatné tri (notebooky Eurostat, Rozloha ľadovca a Mapovacie knižnice) svojím charakterom nevyžadujú funkcionality a nástroje ArcGIS Pro. Zároveň to však nebráni otváraniu, používaniu a spúšťaniu týchto notebookov v ArcGIS Pro rozhraní pre notebooky.

Výsledné notebooky boli vyvíjané tak, aby všetky vstupné dáta boli stiahnuté pomocou Python kódu v rámci notebooku a nebolo nutné prikladať súbor alebo adresár s dátami. Pri dvoch notebookov to však nie je tak: dáta z Urban Atlasu a Sentinelu sú dostupné na stiahnutie až po registrácii a vytvorení účtu na príslušných webových stránkach, a preto je ich sťahovanie pomocou Python kódu veľmi náročné. Vrstva LULC z Urban Atlasu je preto stiahnutá vopred a priložená v geodatabázi zabaleného projektu ArcGIS Pro, ktorý je potrebný na spúšťanie a používanie príslušného notebooku. Satelitné snímky z misie Sentinel-2 je možné sťahovať pomocou knižnice sentinelhub, je však potrebné vložiť prihlasovacie údaje. Tento problém je možné vyriešiť tým, že prihlasovacie údaje by boli zakaždým zadané do užívateľského vstupu pomocou funkcie input. No aj tak je používanie knižnice sentinelhub na sťahovanie snímok komplikované a preto boli potrebné snímky stiahnuté vopred a sú k notebooku priložené v adresári data. Priloženie dát obmedzuje plnú automatizáciu procesu a notebook sa stáva v istom zmysle „časovou kapsulou“, keďže používa dáta platné pre čas, kedy boli stiahnuté. Jedinou výhodou je fakt, že si ďalší užívatelia nemusia vytvárať účet na stiahnutie dát z Urban Atlasu a Sentinelu.

Špecifickým prípadom je notebook ArcPy cvičenia, ktorý pracuje s dátami z ArcCR 500 verzia 3.3, ale tieto dáta nie sú k nemu priložené a ani sa neshahujú pomocou kódu v notebooku. Tento notebook bol vyvinutý primárne pre študentov na KGI, ktorí s dátovou sadou ArcCR pracujú od začiatku štúdia a majú ju stiahnutú tak na školských, ako aj na svojich osobných počítačoch. Čiastočným cieľom cvičení v tomto notebooku je, aby sa študenti naučili správne písať cesty k dátam na lokálnom úložisku, keďže vo výuke im to robilo problém. Keby boli dáta priložené k notebooku, tento cieľ by nebol naplnený. Na druhej strane je veľkosť potenciálnej cieľovej skupiny užívateľov značne limitovaná absenciou vstupných dát.

Varianty spracovania

Variantským spracovaním notebooku Eurostat mohlo byť zapojenie metód pokročilejšej (geo)štatistiky napr. použitím nástrojov z toolboxu ArcGIS Pro Spatial Statistics. Tým by bolo demonštrované efektívne stiahnutie dát z Eurostatu pomocou open-source knižnice a následné spracovanie a analýza týchto dát v ArcGIS Pro. Namiesto zamerania sa na časový vývoj podielu aktívnej populácie vo vede a technológii v európskych regiónoch NUTS 2 medzi rokmi 2011 a 2022 mohol byť notebook zameraný viac na priestorovú distribúciu a koreláciu vysokých a nízkych hodnôt v regiónoch NUTS 2. Eurostat však ponúka viacero štatistík pre európske krajiny a je tu teda priestor na využitie prvej časti notebooku s vyriešeným sťahovaním dát a pokračovaním inej zvolenej priestorovej analýzy.

Notebooky boli vytvorené v programe VS Code. Alternatívnym riešením by bol vývoj notebookov vo webových prostrediach Google Colab alebo JupyterLab a využitie ich hostovaných výpočtových kapacít vrátane rôznych Python kernelov. Medzi ďalšie variantné riešenia patrí využitie iných knižníc na spracovanie a vizualizáciu dát, ako napr. knižnica leafmap. Táto vychádza z knižnice geemap, ktorá je určená na prácu s Google Earth Engine. Leafmap bol vyvinutý pre tých, ktorí nemajú prístup ku Google Earth Engine a umožňuje im analyzovať a vizualizovať geopriestorové dáta s minimom programovania v prostredí Jupyter. Tvorba interaktívnych máp v Leafmap je postavená na knižniciach Folium a ipyleaflet (Wu 2023).

Problémy

Pri vývoji notebookov sa vyskytli dva hlavné problémy. Prvým bolo zistenie nefunkčnosti knižnice SentinelSat na sťahovanie snímok Sentinel-2 z dôvodu prechodu programu Copernicus, ktorý spravuje Sentinel dáta, na nové dátové úložisko [Copernicus Data Space Ecosystem](#). Týmto prechodom sa od 2.11.2023 všetky API a knižnice, ktoré boli naviazané na predchádzajúce úložisko [Copernicus Open Access Hub](#), stávajú nefunkčnými. Tento problém bol vyriešený stiahnutím snímok Sentinel-2 vopred z [Copernicus Browser](#). Druhým problémom bolo zistenie, že notebooky spúšťané v ArcGIS Pro nevedia zobrazíť statickú mapu vytvorenú funkciou plot zo štruktúry GeoDataFrame. Tento problém bol vyriešený tak, že notebooky, ktoré musia byť spúšťané v ArcGIS Pro, používajú na vizualizáciu dát uložených v GeoDataFrame knižnicu Folium. Táto knižnica vytvára interaktívne mapy založené na Leaflet.js a notebook v ArcGIS Pro ich vie bez problémov zobrazíť.

Využitie výsledkov

Štyri z ôsmich výsledných notebookov boli testované vo výuke programovania na KGI, takže ich budúce uplatnenie je vo vzdelávaní študentov. Ostatné štyri notebooky majú tiež potenciál využitia vo výukovom procese, či už programovania alebo v oblastiach ich zamerania. Napríklad notebook Kolokačná analýza by mohol byť použitý vo výuke data miningu alebo modelovania priestorových vzťahov. Notebook Eurostat by mohol byť ukážkou získavania a vizualizácie dát v rámci kurzu o dátových zdrojoch. Notebook Rozloha ľadovca by mohol byť využitý pri vysvetľovaní spektrálneho indexu NDSI na predmete o diaľkovom prieskume Zeme. Notebook Mapovacie knižnice by mohol byť súčasťou výuky kartografie alebo webovej kartografie. Všetky notebooky môžu slúžiť ako prezentácie, cvičenia alebo domáce úlohy.

Výsledky tejto diplomovej práce môžu byť inšpiráciou a návodom pre tých, ktorí pracujú s Jupyter Notebookmi alebo zvažujú ich používanie v geoinformatike. Zverejnením na GitHubu zároveň obohacujú komunitu používateľov Jupyter Notebookov.

Porovnanie notebookov so zdrojmi, ktoré boli ich inšpiráciou

Notebook *Hustota dopravnej siete* vychádza z Python skriptu z bakalárskej práce *Možnosti výpočtu charakteristik dopravní sítě států a měst* (Tóth 2022). Zo skriptu je vytvorený nástroj, ktorý je možné spúšťať v ArcGIS Pro. Dáta dodáva užívateľ pri vyplňaní parametrov nástroja. Tieto dáta kód skriptu kontroluje, či sú validné, aby bola hustota dopravnej siete úspešne a správne vypočítaná. Skript počas doby spustenia komunikuje s užívateľom výpismi o aktuálne vykonávaných činnostiach do okna nástroja. Na konci sú pomocné vrstvy zmazané a užívateľ dostáva jednu výslednú vrstvu s vypočítanými hustotami. Na druhej strane notebook pracuje s dátami, ktoré si „sám“ zabezpečí a stiahne a tým pádom nepotrebuje robiť kontroly dát tak, ako ich robí skript. Postup výpočtu hustoty je rovnaký. Rozdiel je v tom, že pri postupnom spúšťaní buniek notebooku užívateľ vidí a chápe jednotlivé kroky a získa tak oveľa lepšiu predstavu, ako bol výsledok dosiahnutý, než pri spustení skriptu.

Notebook *Eurostat* je podobný článku *Using Eurostat statistical data on Europe with Python* (van der Meulen 2022). Podobnosti spočívajú v prístupe k Eurostat dátam pomocou Python knižnice, v práci s dátami v štruktúre DataFrame, v stiahnutí priestorovej vrstvy z rozhrania GISCO services a vo vizualizácii formou grafov a máp. Koncept je teda podobný, ale spôsob realizácie rozdielny. V článku sú spracované dáta o nezamestnanosti v európskych krajinách, v notebooku dáta o podieli aktívnej populácie vo vede a technológii v európskych regiónoch NUTS 2. Knižnicou na prístup k Eurostat dátam v článku je *EasyEuroStat package*, v notebooku je to knižnica *Eurostat*, ktorá poskytuje oveľa jednoduchší a priamejší prístup ku všetkým Eurostat datasetom. Kým autor článku vytvoril pomocou Python kódu 1 graf a 2 mapy, autor notebooku vytvoril Python kódom 7 grafov a 9 máp. Z článku nebol prevzatý žiaden kód.

Porovnanie notebooku *Rozloha ľadovca* s Python skriptom zo semestrálnej práce *GeoPython: výpočet rozlohy ľadovca* (Potočiar a Tóth 2023) je podobné ako porovnanie notebooku *Hustota dopravnej siete* zo skriptom z uvedenej bakalárskej práce. Jadro kódu ostalo rovnaké, pridaný bol kód, ktorý vytvára vizualizáciu odhadnutej plochy ľadovca pre všetky skúmané roky. Ďalšou zmenou bolo spresnenie prahovej hodnoty NDSI, ktorá určuje hranicu medzi ľadom a ostatnými povrchmi a vytvorenie funkcie na orezávanie snímok. Polygón ohraničujúci ľadovec bol ku skriptu priložený, v notebooku je však stiahnutý knižnicou *OSMnx* priamo v kóde a následne preprojektovaný do súradnicového systému snímok. Pásma snímok sú v skripte ukladané do jedného zoznamu, kým v notebooku sú ukladané do dvoch samostatných zoznamov. Stiahnuté snímky sú priložené aj ku skriptu aj k notebooku, na začiatku notebooku sa nachádza aj návod, ako si ich stiahnuť.

Notebook *Mapovacie knižnice* vychádza z článku *6 python libraries to make beautiful maps* (Rožanov 2024). Koncept je rovnaký, článok aj notebook predstavujú knižnice, ktoré vedia vytvárať mapy. Článok sa venuje šiestim knižniciam, notebook štyrom. Notebook však predstavuje knižnice podrobnejšie (viac príkladov a praktických ukážok) a zároveň ich porovnáva tým, že vizualizuje rovnaké dáta vo všetkých knižniciach. Týmito dátami je polygónová vrstva európskych krajín stiahnutá na začiatku notebooku pomocou Python kódu. Praktických ukážok, ktoré nie sú spoločné, je šesť; ukážky, ktoré majú spoločný základ, sú tri a iba jedna je takmer identická.

Dôraz pri tejto diplomovej práci bol kladený na detailné vypracovanie notebookov. Samotný text práce nie je duplicitným vysvetľovaním postupu v notebookoch, skôr ide o stručnú ilustráciu ich obsahu a popisom motivácie ich cieľov a zvolených riešení.

7 ZÁVER

Cieľom tejto diplomovej práce bolo vytvoriť minimálne 4 ArcGIS Notebooky, ktoré spracovávajú dáta európskych štátov a miest. Každý notebook mal byť zameraný na jednu z vybraných tém: využitie územia v mestskom prostredí, hustota dopravnej siete, automatizovaný prístup k dátam z Eurostatu, spracovanie satelitných snímok a tvorbu máp pomocou Python knižníc. Notebooky majú plniť funkciu „data storytelling“, vďaka ktorému užívateľ získa komplexný prehľad o vybraných témach a spracovaných dátach pomocou postupného spúšťania programového kódu, čítania vysvetľujúceho textu a vizualizácií vo forme tabuliek, grafov a máp. Programovacím jazykom mal byť Python a jazykom, v ktorom majú byť notebooky napísané, mal byť anglický jazyk.

Na vývoj notebookov bol použitý program VS Code, ktorý poskytuje vhodné rozhranie na tvorbu notebookov, editáciu kódu a zálohovanie na GitHub. Kernelom, ktorý spúšťal kód notebookov, bolo naklonované ArcGIS Python prostredie 3.9.18. Do tohto prostredia boli doinštalované potrebné knižnice. Notebooky boli spúšťané a testované v ArcGIS Pro verzia 3.0.2 a aj vo VS Code.

Notebook **Základy Jupyter Notebooku** je úvodom do vytvárania a používania notebookov. Obsahuje základné informácie o tomto programovacom dokumente a jazyku Python a tiež návody, ako formátovať text v markdown bunkách či ako pracovať s premennými v bunkách kódu. Notebook **Python cvičenia** obsahuje osem cvičení, ktoré trénujú programovacie zručnosti užívateľa a v ktorých sa uplatňujú Python príkazy a funkcie uvedené na konci predchádzajúceho notebooku. Notebook **ArcPy cvičenia** je úvodom do používania Jupyter Notebookov v ArcGIS Pro. Obsahuje návod na naklonovanie ArcGIS Python prostredia a osem cvičení, ktoré trénujú používanie základných ArcPy funkcií.

Notebook **Kolokačná analýza** ponúka komplexný prehľad a vysvetlenie kolokačnej analýzy na príklade výskytu mestskej zástavby v blízkosti zelene a športovísk v konurbácii Katowice. Užívateľ má možnosť vidieť, aký vplyv na výsledky analýzy majú jednotlivé nastavenia. Zdrojovými dátami je vrstva využitia územia z Urban Atlasu. Notebook **Hustota dopravnej siete** prechádza procesom výpočtu hustoty železničnej siete na plochu a na počet obyvateľov v meste Liverpool. Vstupnými dátami je železničná sieť z OSM a GEOSTAT populačný grid a obe sú stiahnuté pomocou Python kódu. Notebook **Eurostat** pomocou rovnomennej knižnice pristupuje k dátam z Eurostatu. Vybraný dataset má názov Human resources in science and technology (HRST) by NUTS 2 regions (Eurostat 2024). Tieto dáta sú vizualizované pomocou viacerých grafov a máp, tak statických ako aj interaktívnych.

Notebook **Rozloha ľadovca** predstavuje metódu na odhad veľkosti horských ľadovcov. Vstupnými dátami sú satelitné snímky Sentinel-2 ľadovca Aletschgletscher vo švajčiarskych Alpách. Z týchto snímok je pomocou indexu NDSI vypočítaná rozloha ľadovca a porovnaná zmena tejto rozlohy počas posledných 6 rokov. Notebook **Mapovacie knižnice** predstavuje a porovnáva štyri open-source Python knižnice Cartopy, Folium, Plotly a ipyleaflet, pomocou ktorých je možné vytvárať statické a interaktívne mapy. Vstupnými dátami je polygónová vrstva európskych krajín.

Štyri z ôsmich notebookov (Základy Jupyter Notebooku, Python cvičenia, ArcPy cvičenia, Hustota dopravnej siete) boli testované vo výuke na predmetoch KGI/PROPY a KGI/PRODA na KGI PrF UP. Zvyšné štyri notebooky majú tiež potenciál nasadenia vo výuke, či už ako podklady pre prednášky alebo ako praktické cvičenia. Hlavným prínosom je prezentácia notebookov ako forma interaktívneho programovania, práce s geografickými dátami a ako dokumentov, ktoré rozprávajú príbeh o dátach. Ciele práce boli splnené.

POUŽITÁ LITERATÚRA A INFORMAČNÉ ZDROJE

- ARCDATA PRAHA, ZÚ a ČSÚ, 2016. *ArcČR 500 verze 3.3 – Popis dat* [online]. Dostupné na: https://download.arcdata.cz/data/ArcCR_500-3.3-Popis-dat.pdf
- BATISTA E SILVA, F, L DIJKSTRA a H POELMAN, 2021. *The JRC-GEOSTAT 2018 population grid. JRC Technical Report*.
- BOEING, Geoff, 2017. *OSMnx: New Methods for Acquiring, Constructing, Analyzing, and Visualizing Complex Street Networks* [online]. 2017. Dostupné na: <https://geoffboeing.com/publications/osmnx-complex-street-networks/>
- CARTOPY CONTRIBUTORS, 2024. *Cartopy documentation* [online]. Dostupné na: <https://scitools.org.uk/cartopy/docs/latest/index.html>
- CAZZANIGA, Noemi Emanuela, 2024. *Eurostat Python Package. PyPI* [online]. Dostupné na: <https://pypi.org/project/eurostat/>
- ČERNÝ, Lukáš, 2024. *Automatizace zpracování geografických dat pomocí ArcGIS Pro a programovacího jazyka Python* [online]. B.m. Česká zemědělská univerzita v Praze. Dostupné na: <https://theses.cz/id/zvif7o/?issnshret=ArcGIS%3B;zpet=%2Fvyhledavani%2F%3Fsearch%3Darcgis%20notebook%26start%3D1>
- COPERNICUS LAND MONITORING SERVICE, 2020. *Urban Atlas Land Cover/Land Use 2018* [online]. Dostupné na: <https://land.copernicus.eu/en/products/urban-atlas/urban-atlas-2018?tab=download>
- COTE, Catherine, 2021. *Data Storytelling: How to Effectively Tell a Story with Data. Harvard Business School Online* [online]. 2021. Dostupné na: <https://online.hbs.edu/blog/post/data-storytelling>
- ESA, no date. *Level-2A Algorithm Overview* [online]. Dostupné na: <https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-2-msi/level-2a/algorithm-overview>
- ESRI, no date. *ArcGIS Pro Python reference* [online]. Dostupné na: <https://pro.arcgis.com/en/pro-app/latest/arcpy/main/arcgis-pro-arcpy-reference.htm>
- ESRI, no date. *Colocation Analysis (Spatial Statistics)* [online]. Dostupné na: <https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/colocationanalysis.htm>
- ESRI, no date. *How Colocation Analysis works* [online]. Dostupné na: <https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/learnmorecolocationanalysis.htm>
- ESRI, no date. *Learn Python with ArcGIS Notebooks* [online]. Dostupné na: <https://learn.arcgis.com/en/paths/learn-python-with-arcgis-notebooks/>
- ESRI, no date. *Notebooks in ArcGIS Pro* [online]. Dostupné na: <https://pro.arcgis.com/en/pro-app/latest/arcpy/get-started/pro-notebooks.htm>
- ESRI - ARCGIS API FOR PYTHON, no date. *Chennai Floods 2015 – A Geographic Analysis* [online]. Dostupné na: <https://developers.arcgis.com/python/samples/chennai-floods-analysis-rn/#natures-fury-or-human-made-disaster>
- EUROPEAN COMMISSION, 2017. *Legal notice on the use of Copernicus Sentinel Data and Service Information* [online]. Dostupné na: https://sentinel.esa.int/documents/247904/690755/Sentinel_Data_Legal_Notice

EUROSTAT, 2024. *Human resources in science and technology (HRST) by NUTS 2 regions* [online]. Dostupné na: <https://ec.europa.eu/eurostat/databrowser/view/tgs00038/default/table?lang=en>

EUROSTAT, no date. *Administrative units / Statistical units* [online]. Dostupné na: <https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units>

EUROSTAT, no date. *GISCO* [online]. Dostupné na: <https://ec.europa.eu/eurostat/web/gisco/overview>

EUROSTAT, no date. *Human Resources in Science & Technology (hrst): Reference Metadata in Euro SDMX Metadata Structure (ESMS)* [online]. Dostupné na: https://ec.europa.eu/eurostat/cache/metadata/en/hrst_esms.htm

GEOPANDAS DEVELOPERS, 2022. *GeoPandas API Reference* [online]. Dostupné na: <https://geopandas.org/en/stable/docs/reference.html>

GILLIES, Sean a SHAPELY CONTRIBUTORS, 2024. *Shapely documentation* [online]. Dostupné na: <https://shapely.readthedocs.io/en/stable/index.html>

GOOGLE, no date. *Google colab* [online]. Dostupné na: <https://colab.google/>

HAPALA, Jan, 2021. Plotting maps with European data in Python: Part I. *Medium* [online]. 2021. Dostupné na: <https://jan-46106.medium.com/plotting-maps-with-european-data-in-python-part-i-decd83837de4>

HASSAN, Abdishakur, 2019. Satellite imagery access and analysis in Python & Jupyter notebooks. *Towards Data Science* [online]. 2019. Dostupné na: <https://towardsdatascience.com/satellite-imagery-access-and-analysis-in-python-jupyter-notebooks-387971ece84b>

IBM DOCUMENTATION, 2021. *Jupyter Notebook runtime environment* [online]. Dostupné na: <https://www.ibm.com/docs/en/db2-event-store/2.0.0?topic=notebooks-jupyter-notebook-runtime-environment>

JUPYTER DEVELOPMENT TEAM, no date. *Introduction to ipyleaflet* [online]. Dostupné na: <https://ipyleaflet.readthedocs.io/en/latest/>

JUPYTER DEVELOPMENT TEAM, no date. *Jupyter Project* [online]. Dostupné na: <https://jupyter.org/>

JUPYTER TEAM, 2015. *Markdown Cells* [online]. Dostupné na: [https://jupyter-notebook.readthedocs.io/en/stable/examples/Notebook/Working With Markdown Cells.html](https://jupyter-notebook.readthedocs.io/en/stable/examples/Notebook/Working%20With%20Markdown%20Cells.html)

JUPYTER TEAM, 2024. *JupyterLab Documentation* [online]. Dostupné na: <https://jupyterlab.readthedocs.io/en/stable/>

MADONNA, Thomas a Shekhar SATYARUPA, 2015. Chennai's Rain Check. *Economic & Political weekly* [online]. 2015. Dostupné na: <https://www.epw.in/journal/2015/49/commentary/chennais-rain-check.html>

MAPBOX, 2018. *Rasterio: access to geospatial raster data* [online]. Dostupné na: <https://rasterio.readthedocs.io/en/stable/>

NIETO, Alberto, 2017. Three Reasons to use Jupyter Notebooks as a GIS User. *ArcGIS Blog* [online]. 2017. Dostupné na: <https://www.esri.com/arcgis-blog/products/analytics/analytics/three-reasons-to-use-jupyter-notebooks-as-a-gis-user/>

NUMPY DEVELOPERS, 2022. *NumPy documentation* [online]. Dostupné na: <https://numpy.org/doc/stable/>

OPENAI, 2023. *ChatGPT 3.5* [online]. Dostupné na: <https://chat.openai.com/>

OPENSTREETMAP CONTRIBUTORS, no date. *OpenStreetMap Copyright and License* [online]. Dostupné na: <https://www.openstreetmap.org/copyright>

PANDAS VIA NUMFOCUS INC., 2024. *pandas API Reference* [online]. Dostupné na: <https://pandas.pydata.org/docs/reference/index.html>

PINARD, Philippe, 2022. *matplotlib-scalebar* [online]. Dostupné na: <https://pypi.org/project/matplotlib-scalebar/>

PLOTLY, 2024. *Plotly Open Source Graphing Library for Python* [online]. Dostupné na: <https://plotly.com/python/>

POTOČIAR, Michal a Adam TÓTH, 2023. *GeoPython: výpočet rozlohy řadovca*.

PROJECT JUPYTER, 2023. *Jupyter Widgets* [online]. Dostupné na: <https://ipywidgets.readthedocs.io/en/stable/>

PYTHON SOFTWARE FOUNDATION, 2024a. *os – Miscellaneous operating system interfaces* [online]. Dostupné na: <https://docs.python.org/3/library/os.html>

PYTHON SOFTWARE FOUNDATION, 2024b. *re – Regular expression operations* [online]. Dostupné na: <https://docs.python.org/3/library/re.html>

PYTHON SOFTWARE FOUNDATION, 2024c. *zipfile – Work with ZIP archives* [online]. Dostupné na: <https://docs.python.org/3/library/zipfile.html>

R CODER, 2023. *Stacked bar charts in Matplotlib* [online]. Dostupné na: <https://python-charts.com/part-whole/stacked-bar-chart-matplotlib/>

REITZ, Kenneth, 2019. *Requests: HTTP for Humans™* [online]. Dostupné na: <https://requests.readthedocs.io/en/latest/>

ROZANOV, Aleksei, 2024. 6 python libraries to make beautiful maps. *Medium* [online]. Dostupné na: <https://medium.com/@alexroz/6-python-libraries-to-make-beautiful-maps-9fb9edb28b27>

ŠRAMO, Benjamín, 2023. *Automation of processing GNSS track records for designing intensity maps* [online]. B.m. Palacký University in Olomouc. Dostupné na: <https://www.geoinformatics.upol.cz/dprace/magisterske/sramo23/index.html>

STORY, Rob, 2024. *Folium documentation* [online]. Dostupné na: <https://python-visualization.github.io/folium/latest/index.html>

SYKAS, Dimitris, 2020. Spectral Indices with multispectral satellite data. *Geo University* [online]. 2020. Dostupné na: <https://www.geo.university/pages/blog?p=spectral-indices-with-multispectral-satellite-data>

TDS EDITORS, 2018. About Towards Data Science. *Towards Data Science* [online]. 2018. Dostupné na: <https://towardsdatascience.com/about-towards-data-science-d691af11cc2f>

THE MATPLOTLIB DEVELOPMENT TEAM, 2024. *Matplotlib API Reference* [online]. Dostupné na: <https://matplotlib.org/stable/api/index>

TÓTH, Adam, 2022. *Možnosti výpočtu charakteristik dopravní sítě států a měst* [online]. B.m. Univerzita Palackého v Olomouci. Dostupné na: <https://www.geoinformatics.upol.cz/dprace/bakalarske/toth22/index.html>

VAN DER MEULEN, Leo, 2022. Using Eurostat statistical data on Europe with Python. *Towards Data Science* [online]. 2022. Dostupné na: <https://towardsdatascience.com/using-eurostat-statistical-data-on-europe-with-python-2d77c9b7b02b>

W3SCHOOLS, 2024. *Python tutorial* [online]. Dostupné na: <https://www.w3schools.com/python/default.asp>

WU, Qiusheng, 2023. *leafmap* [online]. Dostupné na: <https://leafmap.org/>

XARRAY DEVELOPERS, 2024. *Xarray documentation* [online]. Dostupné na:
<https://docs.xarray.dev/en/stable/index.html>

Internetové zdroje dokumentácií sú uvedené priamo v texte.

PRÍLOHY

ZOZNAM PRÍLOH

Voľné prílohy

Príloha 1 Poster

Elektronické prílohy

Príloha 2 Digi médium

Popis štruktúry odovzdávaných digitálnych dát na dátové úložisko katedry (príloha 2 digi médium)

Adresáre:

Text_Prace

Vstupne_Data

0_basics

1_colocation_analysis

2_network_density

3_eurostat

4_glacier_area

5_mapping_libraries

Vystupne_Data

0_basics

1_colocation_analysis

2_network_density

3_eurostat

4_glacier_area

5_mapping_libraries

WEB

externe_subory

obrazky

zdroje

Poster