

Univerzita Palackého v Olomouci
Přírodovědecká fakulta
Katedra geoinformatiky

**ANALÝZA VYTÍŽENOSTI ŽELEZNIČNÍ
INFRASTRUKTURY**

Diplomová práce

Jan CINKE

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

Olomouc 2020
Geoinformatika

ANOTACE

Předkládaná diplomová práce si dává za cíl analyzovat data o vytížení železniční infrastruktury z hlediska času a rozdělení do kategorií. Analyzovaná data jsou unikátním datasetem o železniční infrastruktuře poskytnuta, zcela poprvé pro takovéto zpracování, Správou železnic. Data byla poskytnuta za období 2016 až 2019 včetně a obsahují 2 datasety, jeden za monitorovací body a druhý za monitorovací úseky. V úvodní části se práce věnuje teoretické části, ve které popisuje problematiku z hlediska železnice, legislativy, odborných prací, ale i použitých metod. Praktická část je nejdříve zaměřena na popis a předzpracování poskytnutých dat, v rámci, kterého je kladen důraz na přípravu dat pro následné analýzy. Poté je analyzována časová složka dat formou analýzy časových řad s následnou interpretací. Její obsahem je popis průběhu vybraných monitorovacích úseků za dostupné časové období. Následně je prezentováno několik způsobů možné kategorizace s výsledky a jejich interpretací. Konkrétně jde o pět způsobů kategorizace založených na procentuálním vyjádření typu dopravy a shlukování, to vše na základě různých kombinací atributů dostupných v poskytnutých datech. Celá práce je doplněna grafy a vhodnými kartografickými výstupy.

KLÍČOVÁ SLOVA

železniční infrastruktura; shlukování; analýza časových řad

Počet stran práce: 71

Počet příloh: 10 (z toho 6 tištěných, 2 volné a 2 elektronické)

ANOTATION

The presented diploma thesis aims to analyse the data of the workload on railway infrastructure in terms of time and division into categories. The analysed data is a unique data set of railway infrastructure provided for the first time for processing by the Railway Administration. The data were provided for the period 2016 to 2019 and contain 2 datasets, one for monitoring points and the other for monitoring sections. In the introductory part, the thesis deals with the theoretical part, in which it describes the issues in terms of railways, legislation, scientific work, but also the methods used. The practical part is first focused on the description and pre-processing of the provided data, which emphasizes the preparation of data for subsequent analysis. Then the time component of the data is analysed in the form of time series analysis with subsequent interpretation. Its content is a description of the course of selected monitoring sections for the available time period. Subsequently are presented several ways of possible categorization with results and their interpretation. Specifically, there are five ways of categorization based on the percentage of traffic type and clustering, all based on the different combinations of attributes available in the provided data. The whole work is supplemented by graphs and suitable cartographic outputs.

KEYWORDS

Railway infrastructure; clustering; time series analysis

Number of pages: 71

Number of appendixes: 10

Dílčí části diplomové práce byly realizovány v rámci projektu IGA_PrF_2020_027 „*Pokročilé aplikace geoinformačních technologií pro prostorové analýzy, modelování a vizualizace jevů reálného světa*“ podpořené interní grantovou agenturou Univerzity Palackého v Olomouci.

Prohlašuji, že

- diplomovou práci včetně příloh, jsem vypracoval samostatně a uvedl jsem všechny použité podklady a literaturu.

- jsem si vědom, že na moji diplomovou práci se plně vztahuje zákon č.121/2000 Sb. - autorský zákon, zejména § 35 – využití díla v rámci občanských a náboženských obřadů, v rámci školních představení a využití díla školního a § 60 – školní dílo,

- beru na vědomí, že Univerzita Palackého v Olomouci (dále UP Olomouc) má právo nevýdělečně, ke své vnitřní potřebě, diplomovou práci užívat (§ 35 odst. 3),

- souhlasím, aby jeden výtisk diplomové práce byl uložen v Knihovně UP k prezenčnímu nahlédnutí,

- souhlasím, že údaje o mé diplomové práci budou zveřejněny ve Studijním informačním systému UP,

- v případě zájmu UP Olomouc uzavřu licenční smlouvu s oprávněním užít výsledky a výstupy mé diplomové práce v rozsahu § 12 odst. 4 autorského zákona,

- použít výsledky a výstupy mé diplomové práce nebo poskytnout licenci k jejímu využití mohu jen se souhlasem UP Olomouc, která je oprávněna v takovém případě ode mne požadovat přiměřený příspěvek na úhradu nákladů, které byly UP Olomouc na vytvoření díla vynaloženy (až do jejich skutečné výše).

V Olomouci dne

Jan CINKE

Děkuji vedoucí práce doc. Ing. Zdeně Dobešové, Ph.D. za vedení, podněty a rady při vypracování této práce. Dále děkuji Mgr. Michalu Kučerovi za konzultace při pochopení a zpracování poskytnutých dat. Za poskytnutá data děkuji společnosti Správa železnic státní organizace.

ZADÁNÍ DIPLOMOVÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: **Bc. Jan CINKE**
Osobní číslo: **R18861**
Studijní program: **N1301 Geografie**
Studijní obor: **Geoinformatika**
Název tématu: **Analýza dat vytíženosti železniční infrastruktury**
Zadávající katedra: **Katedra geoinformatiky**

Z á s a d y p r o v y p r a c o v á n í :

Cílem práce je provést časovou analýzu vytíženosti železniční infrastruktury. Student provede vyhodnocení naměřených dat na železniční infrastruktuře v České republice za období roku 2016 a 2017. Pro analýzu využije metody data miningu s cílem provést popis a kategorizaci železničních stanic včetně analýzy časové složky dat.

Student vyplní údaje o všech datových sadách, které vytvořil nebo získal v rámci práce, do Metainformačního systému katedry geoinformatiky a současně vytvoří zálohu údajů ve formě validovaného XML souboru. Celá práce (text, přílohy, výstupy, zdrojová a vytvořená data, XML soubor) se odevzdá v digitální podobě na CD (DVD) a text práce s vybranými přílohami bude odevzdán ve dvou svázaných výtiscích na sekretariát katedry. O diplomové práci student vytvoří webovou stránku v souladu s pravidly dostupnými na stránkách katedry. Práce bude zpracována podle zásad dle Voženílek (2002) a závazné šablony pro diplomové práce na KGI. Povinnou přílohou práce bude poster formátu A2.

Rozsah grafických prací: **dle potřeby**
Rozsah pracovní zprávy: **max. 50 stran**
Forma zpracování diplomové práce: **tištěná**
Seznam odborné literatury:

Hančlová, J., Tvrdý, L.: Úvod do analýzy časových řad, Ekonomická fakulta VŠB-TU Ostrava, 2003
Klímek, P., Kasal R.: Počítačové zpracování data v programu Statistica, 3. díl, Univerzita Tomáše Bati, Zlín, 2007
Šarmanová J.: Metody analýzy dat, VŠB-TU, Ostrava, 2012, ISBN 978-80-248-2565-6.
Petr. P. Metody Data Miningu (část 1, část 2). Pardubice, Univerzita Pardubice. 2014
Shekhar S., Xiong H., Zhou X. Encyclopedia of GIS. Springer, Cham, 2017, ISBN: 978-3-319-17884-4
Voženílek, V. Diplomové práce z geoinformatiky. Univerzita Palackého, Olomouc, 2002, 31 s.

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

Datum zadání diplomové práce: **1. listopadu 2018**

Termín odevzdání diplomové práce: **5. května 2020**

doc. RNDr. Martin Kubala, Ph.D.
děkan

L.S.

UNIVERZITA PALACKÉHO V OLOMOUCI
PŘÍRODOVĚDECKÁ FAKULTA
KATEDRA GEONFORMATIKY
17. listopadu 50, 771 46 Olomouc
-1-

prof. RNDr. Vít Voženílek, CSc.
vedoucí katedry

V Olomouci dne 10. prosince 2018

OBSAH

SEZNAM TABULEK	10
SEZNAM ILUSTRACÍ	11
SEZNAM POUŽITÝCH ZKRATEK	13
ÚVOD	14
1 CÍLE PRÁCE.....	15
2 METODY A POSTUP ZPRACOVÁNÍ.....	16
2.1 Použitá data	16
2.2 Metody zpracování	19
2.3 Použitý software	20
2.4 Postup zpracování.....	21
3 SOUČASNÝ STAV ŘEŠENÉ PROBLEMATIKY	22
3.1 Legislativní rámec vývoje železnice v ČR.....	22
3.2 Správce železniční dopravní cesty	22
3.3 Železniční síť v ČR.....	23
3.4 Práce zabývající se zpracováním dat o infrastruktuře	27
3.5 Shlukování.....	29
3.6 Časové řady.....	30
4 PŘEDZPRACOVÁNÍ DAT	31
4.1 Data pro jednotlivé monitorovací body	31
4.2 Úseková data	33
4.3 Výběr atributů	36
5 ZÁKLADNÍ VIZUALIZACE.....	38
5.1 Mapy	38
5.2 Grafy	42
6 ANALÝZA ČASOVÝCH ŘAD	45
6.1 Kategorie	45
6.2 Úprava dat	46
6.3 Dekompozice časových řad	47
6.4 Hodnocení.....	48
7 KATEGORIZACE STANIC.....	59
7.1 Roční agregace	59
7.2 Měsíční agregace	71
7.3 Hodnocení.....	75

8	VÝSLEDKY	77
9	DISKUZE	79
	ZÁVĚR	82
	POUŽITÁ LITERATURA A INFORMAČNÍ ZDROJE	
	PŘÍLOHY	

SEZNAM TABULEK

Tab. 1 Počet záznamů a úseků poskytnutých dat za úseky	16
Tab. 2 Výčet atributů pro poskytnutá data za úseky	17
Tab. 3 Počet záznamů a monitorovacích bodů pro data za jednotlivé monitorovací body	18
Tab. 4 Výčet atributů pro poskytnutá data za jednotlivé monitorovací body	19
Tab. 5 Základní informace o železniční síti ČR (Statistická ročenka 2019, 2020)	24
Tab. 6 Kategorie procentuální kategorizace poměru počtů a hmotností osobní a nákladní dopravy	60
Tab. 7 Intervaly počtu a hmotnosti osobních a nákladních vlakových souprav pro prvotní pojmenování shluků	64
Tab. 8 Intervaly průměrného počtu vlakových souprav celkem pro finální pojmenování shluků	67
Tab. 9 Intervaly průměrné hmotnosti vlakových souprav celkem pro finální pojmenování shluků	70

SEZNAM ILUSTRACÍ

Obr. 3.1 Tranzitní koridory.....	26
Obr. 4.1 Náhled kontingenční tabulky ročních úhrnů pro data za jednotlivé monitorovací body.	32
Obr. 4.2 Náhled kontingenční tabulky měsíčních úhrnů pro data za jednotlivé monitorovací body.	33
Obr. 4.3 Zjednodušený liniový model před úpravou souřadnic.....	34
Obr. 4.4 Zjednodušený liniový model po úpravě souřadnic.	35
Obr. 4.5 Náhled kontingenční tabulky měsíčních úhrnů pro data za úseky.....	35
Obr. 4.6 Model pro výpočet korelace v programu Orange.	36
Obr. 4.7 Vypočtené korelace všech kombinací vstupních atributů bez rozlišení dopravy.	36
Obr. 4.8 Vypočtené korelace počtu a hmotností osobních a nákladních vlakových souprav.	37
Obr. 5.1 Celkový počet vlakových souprav na úsecích železniční sítě ČR v roce 2019..	39
Obr. 5.2 Celková hmotnost vlakových souprav na úsecích železniční sítě ČR v roce 2019.	40
Obr. 5.3 Procentuální podíl počtu osobních vlakových souprav na celkové počtu v ČR v roce 2019.....	41
Obr. 5.4 Procentuální podíl hmotností osobních vlakových souprav na celkové hmotnosti v ČR v roce 2019.	42
Obr. 5.5 Graf ročního průběhu počtu vlaků ve stanici Praha hl. nádraží v roce 2016..	43
Obr. 5.6 Graf ročního průběhu počtu vlaků ve stanici Břeclav v roce 2016.	43
Obr. 5.7 Graf ročního průběhu počtu vlaků ve stanici Adršpach v roce 2016.	44
Obr. 6.1 Karta Timeseries v programu Orange.	47
Obr. 6.2 Model pro rozklad časových řad v programu Orange.....	48
Obr. 6.3 Počet osobních vlakových souprav na vybraných koridorových úsecích v ČR, 2016–2019.	50
Obr. 6.4 Počet nákladních vlakových souprav na vybraných koridorových úsecích v ČR, 2016–2019.	51
Obr. 6.5 Počet osobních vlakových souprav na vybraných hraničních přechodech v ČR, 2016-2019.....	53
Obr. 6.6 Počet nákladních vlakových souprav na vybraných hraničních přechodech v ČR, 2016-2019.	54
Obr. 6.7 Počet osobních vlakových souprav na zajímavých úsecích v ČR, 2016-2019.	56
Obr. 6.8 Počet nákladních vlakových souprav na zajímavých úsecích v ČR, 2016-2019.	57

Obr. 6.9 Počet osobních a nákladních vlakových souprav v úseku Mladá Boleslav město, 2016-2019.....	58
Obr. 7.1 Schéma rozdělení kategorizací na základě metody a vstupních dat.....	59
Obr. 7.2 Procentuální podíl dopravy z hlediska počtu vlakových souprav v roce 2016.	61
Obr. 7.3 Procentuální podíl dopravy z hlediska hmotnosti vlakových souprav v roce 2016.	62
Obr. 7.4 Model hierarchického shlukování v programu Orange.	63
Obr. 7.5 Parametrizace nástroje Distance v programu Orange.....	63
Obr. 7.6 Parametrizace a dendrogram hierarchického shlukování v programu Orange.	64
Obr. 7.7 Výsledek shlukování ročních úhrnů počtu osobních a nákladních vlaků v roce 2016.	65
Obr. 7.8 Průměrné hodnoty vstupních atributů v jednotlivých shlucích počtu osobní a nákladní dopravy pro rok 2016.	67
Obr. 7.9 Výsledek shlukování ročních úhrnů hmotnosti osobních a nákladních vlaků v roce 2016.....	68
Obr. 7.10 Průměrné hodnoty vstupních atributů v jednotlivých shlucích hmotnosti osobní a nákladní dopravy pro rok 2016.	70
Obr. 7.11 Výsledek shlukování měsíčních úhrnů hmotnosti vlakových souprav roce 2016.	72
Obr. 7.12 Průměrné hodnoty vstupních atributů v jednotlivých shlucích hmotnosti vlakových souprav pro jednotlivé měsíce roku 2016.	73
Obr. 7.13 Méně početný shluk při shlukování relativní dat hmotnosti vlakových souprav o počtu shluků 20.	74
Obr. 7.14 Více početný shluk při shlukování relativní dat hmotnosti vlakových souprav o počtu shluků 20.	75

SEZNAM POUŽITÝCH ZKRATEK

ČSÚ	Český statistický úřad
ČÚZK	Český úřad zeměměřický a katastrální
ČVUT	České vysoké učení technické
Esri	Environmental Systems Research Institute
GPS	Global Positioning System
IGA	Interní grantová agentura
L _{dv}	hlukový indikátor pro celkové obtěžování hlukem
L _n	hlukový indikátor pro rušení spánku
SR70	Číselník železničních stanic, dopravně zajímavých a tarifních míst
SŽ	Správa železnic
SŽDC	Správa železniční dopravní cesty

ÚVOD

V roce 2018 byla na Katedru geoinformatiky, Univerzity Palackého v Olomouci poskytnuta data o vytížení železniční infrastruktury od Správy železnic České republiky (dále SŽ). Tato data původně sloužila pouze jako primární zdroj při výpočtu ceny za využití dráhy a jsou generována na základě smluv s jednotlivými dopravci působícími na železnicích spravovaných SŽ. K uvolnění těchto dat došlo v tomto případě poprvé, právě pro jejich využití na Katedře geoinformatiky v Olomouci. Jde o zcela unikátní dataset informací o provozu na železniční síti ČR, který je vázán na železniční stanice, dopravně zajímavá a tarifní místa, která jsou vedena ve služební rukověti SR 70 vydávané SŽ. Jak je patrné z výše zmíněného, dataset nebyl nikdy využíván na bázi prostorových a neprostorových analýz.

Na základě možnosti, která byla dána spoluprací mezi katedrou geoinformatiky a SŽ, bylo rozhodnuto o zpracování těchto dat, která mají velký potenciál z hlediska analýzy provozu na železniční síti v ČR. Vzhledem k povaze těchto dat, která byla poskytnuta jako měsíční úhrny ze čtyř po sobě jdoucích let, je možné využít nejen možnosti dané popisnými informacemi těchto dat, ale také jejich časovou variabilitu.

Podstatou této diplomové práce je kategorizace poskytnutých dat a analýza na základě jejich časové variability. Data jsou nejdříve předzpracována do podoby, ve které je možné s nimi dále pracovat. Na základě upravených dat je provedena analýza časových řad formou dekompozice z hlediska počtu vlakových souprav za měsíc pro vybrané kategorie železničních úseků. Poté je provedeno několik možností kategorizace těchto dat a jejich vyhodnocení. Interpretační část časové analýzy a kategorizací je doprovázena vhodnými mapovými výstupy a grafy. Pojem vytížení železniční infrastruktury je v této práci používán jako zatížením železniční sítě, zejména z hlediska počtu a hmotnosti vlakových souprav.

1 CÍLE PRÁCE

Cílem práce je provést časovou analýzu využitosti železniční infrastruktury. Na základě poskytnutých dat bude provedeno vyhodnocení dat na železniční infrastruktuře v České republice za období roku 2016 až 2019. V první části budou data předzpracována do vhodné podoby pro další zpracování v této diplomové práci i mimo ni. V další části bude provedena analýza časové složky dat, tak aby bylo možné popsat časovou variabilitu vybraných úseků v rámci ČR. Poté bude provedeno několik kategorizací bodových vstupních dat, na základě rozdílných metod a atributů, které budou popsány. Kategorizace a analýza časových řad budou doprovázeny vhodnými mapovými výstupy a grafy.

2 METODY A POSTUP ZPRACOVÁNÍ

2.1 Použitá data

Klíčovou částí této diplomové práce jsou data o vytížení železniční infrastruktury poskytnutá státní organizací SŽ ve formátu XLSX. Data se používají jako primární zdroj při výpočtu ceny za využití dráhy jednotlivými dopravci. Data jsou reprezentována měsíčními úhrny informací o vytížení za jednotlivé druhy kolejové dopravy. Jsou rozdělena na dvě části s odlišnou strukturou. První částí jsou data za jednotlivé úseky mezi monitorovacími body, druhým typem dat jsou úhrny za jednotlivé monitorovací body.

2.1.1 Úseková data

Jde o měsíční úhrny informací o využití železniční infrastruktury za úseky mezi monitorovacími body. Tyto data byla postupně poskytována během zpracování práce s výsledným rozsahem od roku 2016 do roku 2019 včetně. Jejich struktura je sestavena tak, že každý řádek obsahuje informace o vytížení určitého úseku reprezentovaného jeho počátečním a koncovým bodem. Každý úsek je v datech pouze jednou, nejde tedy o data směrová, ale o úhrny provozu mezi dvěma body v obou směrech. Výčet záznamů a úseků je dostupný níže (tab. 1).

Tab. 1 Počet záznamů a úseků poskytnutých dat za úseky

rok	počet záznamů	počet úseků
2016	241 706	3 782
2017	248 141	3 777
2018	253 408	3 873
2019	260 740	3 542

Atributy se dají rozdělit do čtyř částí. V první části je označení období, tedy měsíce a roku ze kterého datový záznam pochází. V druhé části jsou popisné informace úseku, jde o informace o poloze, kódovém označení a názvu počátečního a koncového bodu. Ve třetí části jsou obsaženy informace o kategorii dopravy daného záznamu. V poslední části jsou samotná číselná data o využití železniční infrastruktury. Některé atributy jsou absolutní za celý měsíc, některé průměrné. Kompletní výpis atributů je zobrazen níže (tab. 2).

Tab. 2 Výčet atributů pro poskytnutá data za úseky

atribut	popis
datum numericky	datum, ke kterému jsou záznamy vztaheny ve formátu RRRRMMDD
číslo úseku dle KKM	interní označení jednotlivých úseků
výchozí dopravní bod SR70	jedinečné šestimístné číslo výchozího monitorovacího bodu určené služebním předpisem SR70
název výchozího dopravního bodu	slovní označení výchozího bodu
GPS souřadnice výchozího dopravního bodu	GPS souřadnice výchozího bodu ve formátu ESS°MM'VV" - NSS°MM'VV"
SR70 koncového dopravního bodu	jedinečné šestimístné číslo koncového monitorovacího bodu určené služebním předpisem SR70
název koncového dopravního bodu	slovní označení koncového dopravního bodu
GPS souřadnice koncového dopravního bodu	GPS souřadnice cílového bodu ve formátu ESS°MM'VV" - NSS°MM'VV"
druh vlakové soupravy	podrobné dělení osobních, nákladních a služebních vlakových souprav pomocí zkratk definovaných předpisem SŽ
kód kategorie dopravy	označení dělení kategorie dopravy číselným kódem: 1 – osobní, 2 – nákladní, 3 – služební, 4 – lokomotivní, 5 – soupravový
kategorie dopravy	dělení na základní kategorie vlakových souprav jako jsou: osobní, nákladní, služební, lokomotivní a soupravový
počet vlakových souprav	měsíční úhrn počtu vlakových souprav
průměrná délka soupravy	průměrná délka soupravy za měsíc
hmotnost souprav	celková hmotnost souprav za měsíc
průměrná hmotnost soupravy	průměrná hmotnost souprav za měsíc
průměrný počet náprav	průměrný počet náprav za měsíc
průměrný počet vozů	průměrný počet vozů za měsíc

Při plánovaném zpracování těchto dat z hlediska prostoru se vyskytl problém chybějící liniové síťové vrstvy, ke které by se tato data dala připojit a mohl by tak být vybudován síťový model s informacemi o vytižení železniční infrastruktury. Důvod je takový, že dělení těchto dat na úseky je zcela specifické. Dalším problémem, tentokrát z hlediska kategorizace na základě číselných atributů byla velká korelace mezi jednotlivými atributy. Například váha soupravy koreluje s její délkou, nebo počet vagonů koreluje s počtem náprav. V případě zpracování dat z hlediska osobní dopravy vznikaly dva problémy. Za prvé je důležité si uvědomit, že data neposkytují informace o vlacích, které obsluhují úsek s danou zastávkou, ale informace o všech osobních a nákladních vlacích, které projeli daným úsekem. Druhým problémem je, že je komplikované na základě dat jednoznačně rozlišit, které počáteční a koncové monitorovací body jsou stanicemi (z pohledu osobní dopravy), a které hradly, výhybkami, nebo jinými tarifně důležitými místy. Nelze tak provést ani srovnáním s jinou vrstvou železniční stanic (např. OSM, nebo Data200), protože samotné body mají mírně odlišnou polohu, nebo název.

2.1.2 Data za jednotlivé monitorovací body

Jde o měsíční úhrny informací o využití železniční infrastruktury za jednotlivé monitorovací body. Data byla poskytnuta jako celek v rozsahu kompletního roku 2016 a prvních 10 měsíců roku 2017. Jejich struktura je od úsekových dat odlišná, zejména v samotné povaze monitorované jednotky. Zde nahrazují úseky mezi monitorovacími body úhrny za jednotlivé monitorovací body. Jeden záznam v tabulce tedy reprezentuje měsíční úhrn informací za jeden měřicí bod s jedním typem dopravy. Výčet záznamů a monitorovacích bodů je dostupný níže (tab. 3).

Tab. 3 Počet záznamů a monitorovacích bodů pro data za jednotlivé monitorovací body

rok	počet záznamů	počet monitorovacích bodů
2016	592 981	3 527
prvních 10 měsíců 2017	455 420	3 488

Můžeme zde opět použít rozdělení na části. První částí je stejně jako u předchozího datasetu, jde tedy o označení měsíce a roku ze kterého záznam pochází. Druhá část popisuje druh vlakové soupravy. Ve třetí části jsou popisné informace o monitorovacím bodu. Ve čtvrté části už jsou stejně jako v předchozím případě číselná data využití železniční infrastruktury. Opět jsou zde některé atributy absolutní za celý měsíc, některé průměrné. Kompletní výpis atributů je zobrazen níže (tab. 4).

Tab. 4 Výčet atributů pro poskytnutá data za jednotlivé monitorovací body

atribut	popis
období numericky	datum, ke kterému jsou záznamy vztaheny ve formátu RRRRMMDD
druh vlakové soupravy	označení jednotlivých druhů vlakových souprav zkratkou
příznak, jestli je vlak nákladní	binární označení nákladní vlakové soupravy: nákladní-1, není nákladní-2
SR70	jedinečné šestimístné číslo monitorovacího bodu určené služebním předpisem SR70
název dopravního bodu	slovní označení dopravního bodu
název dopravního bodu (35 znaků)	slovní označení dopravního bodu zkrácené na 35 znaků, pokud je potřeba
kvalifikátor	slovní označení typu monitorovacího místa, 35 typů
počet vlakových souprav	měsíční úhrn počtu vlakových souprav
hmotnost soupravy	celková hmotnost souprav za měsíc
počet vozů soupravy	počet vozů souprav za měsíc
průměrná délka soupravy	průměrná délka souprav za měsíc
průměrná počet náprav	průměrný počet náprav za měsíc

Z hlediska prostorového zpracování je zde opět problém s chybějící prostorovou vrstvou, ke které by mohly být data připojena. Bohužel zde ani nejsou dostupné souřadnice jednotlivých monitorovacích bodů. Opět je zde také problém s korelací jednotlivých číselných atributů (viz kap. 2.1.1). Z hlediska zpracování dat jako časové řady je komplikací také velmi krátké období, pouze 22 měsíců. Pokud jde o porovnání let mezi sebou, tak je zde zcela nemožné. Lze pouze porovnat 10 měsíců navzájem.

2.2 Metody zpracování

2.2.1 Kontingenční tabulky

„Kontingenční tabulka je výkonný nástroj pro výpočet, sumarizaci a analýzu dat, který umožňuje zobrazit porovnání, vzorky a trendy v datech“ (Microsoft, 2019). Při použití v této diplomové práci bylo využito řešení kontingenčních tabulek v programu Microsoft Office Excel. Poskytnutá data musela být sumarizována za jednotlivé úseky nebo monitorovací body. Původní stav zobrazoval měsíční záznamy za jednotlivé druhy dopravy pro úseky nebo monitorovací body. Nejdříve byla vytvořena prázdná kontingenční tabulka na novém listu aplikace Excel a poté jsou nadefinovány jednotlivé oblasti tabulky do čtyř kategorií. Kategorie filtry je volitelná a můžeme zde umístit

sloupec, dle kterého chceme výslednou tabulku filtrovat. Řádky je kategorie, do které umístíme sloupec, který bude prvním sloupcem nové tabulky, dle kterého se obsah tabulky přepočítává. V kategoriích sloupce a hodnoty se definuje oblast zdrojového grafu, ze kterého budou pocházet hodnoty nového grafu, sumarizované na základě dostupných funkcí, jako jsou například suma nebo průměr.

2.2.2 Časové řady

„Časovou řadou rozumíme posloupnost hodnot ukazatelů, měřených v určitých časových intervalech. Tyto intervaly jsou zpravidla rovnoměrné (ekvidistantní), a proto je můžeme zapsat následujícím způsobem: y_1, y_2, \dots, y_n neboli $y_t, t = 1, \dots, n$, kde y značí analyzovaný ukazatel, t je časová proměnná s celkovým počtem pozorování n “ (Hančlová a Tvrđý 2003, s. 3). Cílem analyzování časových řad je konstrukce takového modelu, který by pomohl porozumět mechanismu, na jehož základech vznikají hodnoty časových řad. Dále pak pomůže k pochopení podmínek a vazeb, které působí na vznik těchto hodnot. Na základě vybudované modelu, lze predikovat a hodnotit budoucí chování (Hančlová a Tvrđý, 2003). Dle Klímka a Kasala (2007) se časovou řadou rozumí posloupnost věcně a prostorově srovnatelných pozorování, která jsou uspořádána z hlediska času ve směru od minulosti do budoucnosti.

2.2.3 Shlukování

Nejdříve je nutné si vymezit rozdíl mezi shlukováním a klasifikací. Klasifikace je takové dělení objektů, kdy máme předem zadané kategorie, do kterých se budou objekty přiřazovat, nebo je jasně definován postup tohoto rozdělení. Shlukovací metoda je taková, kdy neznáme klasifikační kritéria a současně ani neznáme vlastnosti budoucích shluků. Výsledné shluky můžeme nazývat jako hledané klasifikační třídy (Šarmanová, 2015). „Metody shlukové analýzy mají za úkol hledat podobnost nebo naopak odlišnost mezi objekty jen na základě zadaných vlastností, bez jejich dalšího vysvětlování. Shluková analýza hledá struktury v zpracovávaných datech. Seskupuje podobné objekty do množin, které nazýváme shluky. Prvky jednoho shluku si jsou navzájem nějakým způsobem podobnější, než objekty do tohoto shluku nepatřící“ (Kučera 2008, s. 7). „Shluková analýza zkoumá, zda se množina objektů $O = \{O_1, O_2, \dots, O_m\}$ zadaných reálnými atributy $A = \{A_1, A_2, \dots, A_n\}$ přirozeně rozpadá na výrazné podmnožiny objektů si podobných, a přitom nepodobných objektům shluků ostatních. Pokud takové podmnožiny existují, nazýváme je shluky“ (Šarmanová 2015, s. 65). Samotná míra podobnosti nemá platnou definici stejně jako samotný shluk. Pro výběr shlukovací metody je třeba zvážit velké množství možností zpracování a vstupních atributů.

2.3 Použitý software

Pro zpracování poskytnutých dat byl použit nástroj Microsoft Excel z balíčku Microsoft Office 365 ProPlus. Data byla nejdříve zkontrolována a poté transformována pomocí kontingenční tabulky. Byly zde také vyhotoveny grafy četností dat a časové řady. Pro shlukování a následnou kategorizaci dat byl použit software Orange 3.24.1. Pro vyhotovení prostorové síťové vrstvy a vizualizací byl použit ArcGIS Pro 2.5.0.

2.4 Postup zpracování

Postup při zpracování této práce se dá rozdělit do dvou kroků. Prvním je teoretická část, ve které je provedeno seznámení s tématem, legislativou, metodami a odbornými pracemi. Druhá část je praktická, ve které budou naplněny vhodnou formou vytyčené cíle.

- **Teoretická část:**

- legislativa týkající se železnice v ČR,
- seznámení s železniční sítí v ČR a jejím správcem,
- seznámení se s odbornými a akademickými pracemi zabývajícími se podobnými tématy,
- seznámení se s problematikou shlukováním a časových řad.

- **Praktická část:**

- předzpracování dat – základní seznámení s daty a zpracování do podoby vhodné pro další analýzy,
- tvorba základních vizualizací – tvorba vizuálních pomůcek pro pochopení problematiky, kterou data obsahují,
- kategorizace pomocí shlukování – tvorba několika kategorizací železničních dat na základě různé parametrizace,
- analýza časových řad – popis časové složky dat pomocí rozkladu časové řady na komponenty,
- tvorba grafů a mapových výstupů v prostředí GIS – zobrazení výsledků pomocí grafů a vhodných kartografických metod,
- tvorba posteru a webové stránky.

3 SOUČASNÝ STAV ŘEŠENÉ PROBLEMATIKY

Dle Brinkeho (1999) se doprava obecně nejčastěji definuje jako záměrné a organizované přemístění věcí a osob uskutečňované dopravními prostředky po dopravních cestách. Doprava je dělena na složky, jimiž jsou dopravní cesty, dopravní prostředky a ostatní dopravní infrastruktura. Další dělení je na druhy dopravy, které jsou vzdušná, mořská a pevninská doprava. Ta se dále dělí na silniční, železniční, vnitrozemskou plavbu, potrubní a ostatní. Dle Mirvalda (2000) se železnice v rámci nákladní dopravy podílí na objemu přepravy na střední i větší vzdálenosti, kde konkuruje silniční dopravě. V osobní přepravě je železnice perspektivní jako součást integrovaných dopravních systémů, nebo jako páteřní síť v dálkové dopravě, kde konkuruje nejen silniční přepravě, ale v případě vysokorychlostní železnice, také přepravě letecké.

3.1 Legislativní rámec vývoje železnice v ČR

České dráhy byly zřízeny *zákonem České národní rady č. 9/1992 Sb. ze dne 20. prosince 1992 o Českých dráhách*. Tento zákon nabyl účinnosti k 1. lednu 1993 v souladu se vznikem samostatné České a Slovenské republiky. České dráhy se staly nástupnickým subjektem Československých státních drah, které se zabývaly správou železniční infrastruktury a provozováním železniční dopravy na území celého tehdejšího Československa. Dále byl původní zákon rozšiřován a upravován *zákonem České národní rady č. 212/1993 Sb. ze dne 9. července 1993 a zákonem České národní rady č. 218/1993 Sb. dne 13. srpna 1993*. Rozdělení na dráhy celostátní a regionální proběhlo prostřednictvím *Usnesením vlády České republiky č. 766 ze dne 20. prosince 1995*. Nejdůležitějším organizačním krokem novodobé historie se stalo rozdělení státní organizace České dráhy na dva nástupnické subjekty, kterými jsou akciová společnost České dráhy a státní organizace Správa železniční a dopravní cesty. Transformace proběhla na základě *zákona č. 77/2002 Sb. o akciové společnosti České dráhy, státní organizaci Správa železniční dopravní cesty a o změně zákona č. 266/1994 Sb., o dráhách, ve znění pozdějších předpisů, a zákona č. 77/1997 Sb., o státním podniku, ve znění pozdějších předpisů*. Vznik těchto subjektů je pak datován k 1. lednu 2003, kdy zákon nabyl platnosti a současně s tím pozbyly platnosti *zákony č. 9/1993 Sb. i 212/1993 Sb.* (Historie železnice v ČR, 2019). Na základě zákona, kterým se mění *zákon č. 266/1994 Sb., o dráhách, ve znění pozdějších předpisů, a další související zákony* se s platností od 1. ledna 2020 změnil název organizace Správa železniční dopravní cesty (SŽDC) na Správa železnic (SŽ). Přípustné označení této organizace je nadále možné pouze ve tvaru Správa železnic, státní organizace, nebo Správa železnic (Tiskové prohlášení SŽ, 2020).

3.2 Správce železniční dopravní cesty

SŽ je státní organizace hospodařící s majetkem státu, který tvoří především železniční dopravní cesta. V rámci povinností plní funkci vlastníka železniční infrastruktury, zajišťuje provozování, provozuschopnost, modernizaci a rozvoj železniční dopravní cesty. Dále přerozděluje kapacitu dopravní cesty a provozuje celostátní železniční dráhu a regionální dráhu ve vlastnictví státu (O nás, 2020).

Široký a kol. (2018) se ve své publikaci *Technologie dopravy*, zabývá definicí vlastníka dráhy, provozovatele dráhy a drážní dopravy. Podle autora je vlastníkem dráhy „*subjekt, který vlastní určitý úsek dopravní cesty určený k pohybu drážních vozidel, jehož součástí*

jsou pevná zařízení potřebná pro pohyb drážních vozidel a pro zajištění bezpečnosti a plynulosti drážní dopravy. Je povinen udržovat tuto dráhu tak, aby byla zajištěna její provozuschopnost a umožněn styk s jinou dráhou. Zajištěním provozuschopnosti dráhy se rozumí zabezpečovat takový technický stav, který zaručuje bezpečně a plynulé provozování dráhy“ (Široký a kol. 2018, s. 69). Provozovatelem dráhy je dle autora ten, kdo „provozuje dráhu na základě úředního povolení vydaného Drážním úřadem. Zajišťuje provozování dráhy, tj. činnosti, kterými se zabezpečuje a obsluhuje dráha a organizuje drážní doprava. Je povinen provozovat dráhu podle pravidel pro provozování dráhy, které jsou obsaženy ve vyhl. 173/1995 Sb., kterou vydává Dopravní úřad“ (Široký a kol. 2018, s. 69). Provozovatel drážní dopravy neboli dopravce, dle autora „provozuje drážní dopravu na základě licence udělené Drážním úřadem“ (Široký a kol. 2018, s. 70).

V ČR je většinovým vlastníkem kolejové infrastruktury stát, prostřednictvím státní organizace SŽ. Největšími provozovateli drážní dopravy v oblasti osobní dopravy u nás jsou ČD a. s., RegioJet a. s., LEO Express a. s., Die Länderbahn, ARRIVA vlakové soupravy nebo GW TrainRegio a. s. V nákladní dopravě jsou největšími provozovateli ČD Cargo a. s., AWT a. s., METRANS Rail s.r.o a Unipetrol Doprava s.r.o. (Široký a kol., 2018).

3.3 Železniční síť v ČR

Železniční síť reprezentuje souhrn všech železniční dopravních cest na určitém území, kterým může být kontinent, stát, nebo pouze část státu. Dále pak dělíme železniční síť na menší celky (tratě) oddělené zpravidla významnými železničními stanicemi nebo železničními uzly. Tyto tratě slouží veřejné železniční dopravě. Železniční síť může být rozlišována dle správce železniční infrastruktury. Železniční tratě v jedné železniční síti nemusí mít stejný rozchod, nebo nemusí být elektrifikované jednou proudovou soustavou. V závislosti na počtu kolejí můžeme tratě dělit na jednokolejné, dvoukolejné, trojkolejné a čtyřkolejné, případně vícekolejné. Další dělením vychází ze šířky neboli rozchodu, železniční trati. Základní dělení je na tratě s normálním, širokým a úzkým rozchodem. V závislosti na použití energie jako hnací síly v železniční síti rozlišuje tratě na elektrifikované a neelektrifikované. Tratě vycházející z vlastního území vedené přes území jiného státu a vracující se zpět na území vlastního státu nazýváme peážní tratě (Gašpařík a Kolář, 2017).

Celková délka tratí v ČR k 31. 12. 2019 činí 9 396 km (Statistická ročenka 2019, 2020). Rozloha ČR je 78 865 km² (ČSÚ, 2011). Průměrná hustota je přibližně 0, 12 km tratí na 1 km², což znamená, že ČR má společně s Německem a Belgií nejhustší železniční síť na světě. V převážné většině ČR a zhruba 60 % světa se využívá normálního rozchodu, který činí 1435 mm. Nejpoužívanější trakční soustava v ČR je stejnosměrná trakční soustava 3 kV a střídavá trakční soustava 25 kV/50 Hz (Gašpařík a Kolář, 2017). Základní informace o železniční síti v ČR jsou dostupné níže (tab. 5).

Tab. 5 Základní informace o železniční síti ČR (Statistická ročenka 2019, 2020)

typ trati	délka [km]
jednokolejné	7 372
dvoukolejné	1 965
vícekolejné	59
normální rozchod	9 373
úzký rozchod	23
elektrizované	3 213
neelektrizované	6 183
elektrizované AC 25000 V / 50 Hz	1 372
elektrizované AC 15000 V / 16,7 Hz	14
elektrizované DC 3000 V (1500 V)	1 803
elektrizované DC 1500 V	24

Dle zákona č. 266/1994 Sb., o dráhách jsou železniční tratě kategorizovány z hlediska významu, účelu a technických podmínek takto:

- **dráha celostátní**, již je dráha, která slouží mezinárodní a celostátní veřejné železniční dopravě a je jako taková označena,
- **dráha regionální**, již je dráha regionálního nebo místního významu, která slouží veřejné železniční dopravě a je zaústěná do celostátní nebo jiné regionální dráhy,
- **dráha místní**, již je dráha místního významu oddělená od celostátní nebo regionální dráhy; dráha je oddělená, umožňuje-li přesun drážního vozidla na jinou dráhu jen s použitím zvláštního technického zařízení nebo slouží-li výhradně provozování neveřejné osobní drážní dopravy, osobní drážní dopravy pro potřeby cestovního ruchu nebo provozované historickými vlaky,
- **vlečka**, již je dráha, která slouží vlastní potřebě provozovatele nebo jiného podnikatele a je zaústěná do celostátní nebo regionální dráhy, nebo jiné vlečky,
- **zkušební dráha**, již je dráha, která slouží zejména k provádění zkušebního provozu drážních vozidel nebo zkoušek pro schválení typu nebo změny typu drážních vozidel a drážní infrastruktury,
- **speciální dráha**, která slouží zejména k zabezpečení dopravní obslužnosti obce.

Podmínky pro zařazení do kategorie celostátní dráha jsou následující. V první řadě jde o tratě zařazené do evropského železničního systému kategorie E, nebo systémů AGC a AGTC. Dalším kritériem je intenzita nákladní dopravy, která je měřena v tunách za rok. Tratě, u kterých byl naměřen objem nákladní dopravy vyšší než 900 tis. t/rok spadají do kategorie celostátní dráha. Tratě, které mají hodnotu 150-900 tis. t/rok a méně než 150 tis. t/rok se posuzují podle dalších kritérií. Tratě se 150-900 tis. t/rok musejí mít takt dálkové dopravy alespoň 240 minut. U trati se 150 tis. t/rok je nutné

mít takt dálkové nákladní dopravy minimálně 120 minut. Pokud nedojde ke splnění předchozích kritérií je dráha označena jako regionální. Nehledě na předchozí kritéria je celostátní drahou ta, která zajišťuje hlavní spojení krajských měst, je spojkou dvou významných tratí. Další možností je, že trať prochází městem, kde je významná železniční doprava vedena po obchvatu či v případě, že trať může sloužit jako obchvat. Pokud je dráha významná z hlediska přeshraniční dopravy je možnost ji též považovat za dráhu celostátní. Závěrečným kritériem je návaznost dráhy na linku dálkové dopravy osobním vlakem ve stejném taktu. Pokud není splněno ani jedno z předchozích kritérií pro dráhu celostátní, tak je dráha zařazena do kategorie regionální (Komentář ke studii „Kritéria kategorizace železniční sítě“, 2019).

Vlečka je dráha sloužící k vlastní potřebě provozovatele nebo jiného podnikatele a je zaústěná do celostátní nebo regionální dráhy, nebo jiné vlečky. Speciální dráha, která slouží zejména k zabezpečení dopravní obslužnosti obce (Informace o kategorizaci železniční sítě, 2019).

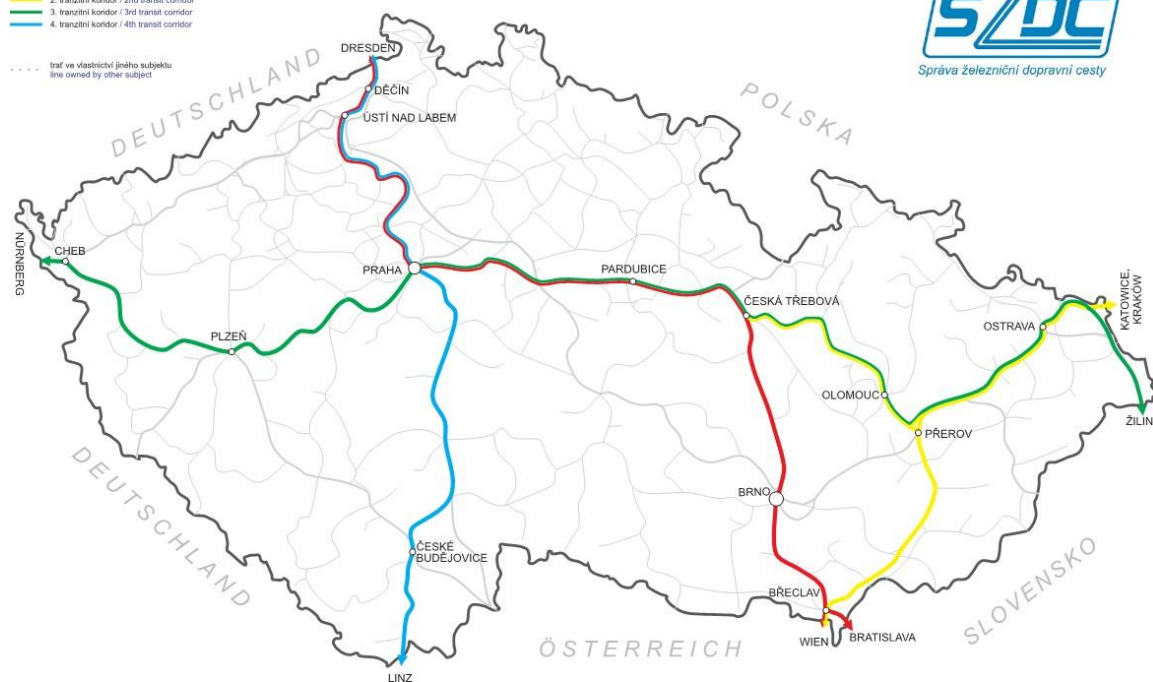
Základem železniční sítě v ČR jsou čtyři hlavní železniční koridory o délce 1329 km (Gašpařík a Kolář, 2017). Tyto tranzitní železniční koridory byly v ČR vymezeny na základě dohod AGC – o mezinárodních železničních magistralách a AGTC – o nejdůležitějších trasách mezinárodní kombinované dopravy a souvisejících objektech. K těmto dohodám ČR přistoupila v rámci plánů rozvoje železničních sítí vypracovaných na úrovni Evropské unie a Mezinárodní železniční unie (UIC) (Tranzitní železniční koridory, 2019). Vymezení tranzitních koridorů dle Tranzitní železniční koridory (2019) v závislosti na evropském rozdělení je následující:

- 1. tranzitní železniční koridor (Východo – středomořský koridor)
 - (Berlín – Drážďany) – **Děčín – Praha – Pardubice – Česká Třebová – Brno – Břeclav** – (Vídeň/ Bratislava – Budapešť),
- 2. tranzitní železniční koridor (Baltsko – jadranský)
 - (Gdaňsk – Varšava – Katowice) – **Petrovice u Karviné – Ostrava – Přerov – Břeclav**,
- 3. tranzitní železniční koridor (Rýnsko – Dunajský)
 - (LeHavre – Paříž – Frankfurt n. M.) – **Cheb – Plzeň – Praha – Ostrava** – (Žilina – Košice – Lvov); odbočná větev Plzeň – Domažlice – (Nürnberg),
- 4. tranzitní železniční koridor
 - (Stockholm – Dresden) – **Děčín – Praha – Tábor – Veselí nad Lužnicí – České Budějovice – Horní Dvořiště** – (Linec – Salzburg – Lublaň – Rijeka – Záhřeb).

Tranzitní koridory / Railway transit corridors

- 1. tranzitní koridor / 1st transit corridor
- 2. tranzitní koridor / 2nd transit corridor
- 3. tranzitní koridor / 3rd transit corridor
- 4. tranzitní koridor / 4th transit corridor

..... trať ve vlastnictví jiného subjektu
line owned by other subject



Obr. 3.1 Tranzitní koridory (Zdroj: <https://bit.ly/2TcSkgj>).

Hlavní tranzitní koridory v ČR (viz obr. 3.1) jsou doplněny o další významné tahy a spojnice Přerov – Brno, Brno – Jihlava – Veselí nad Lužnicí, České Budějovice – Plzeň, Cheb – Chomutov – Ústí nad Labem, Děčín – Nymburk – Kolín – Havlíčkův Brod – Brno, Hranice na Moravě – Horní Lideč, Brno – Veselí nad Moravou – Vlárský průsmyk. Na základě *nařízení Evropského parlamentu a Rady (EU) č. 913/2012 o evropské železniční síti pro konkurenceschopnou nákladní dopravu* vznikly pro nákladní dopravu mezinárodní železniční koridory (RFC). Podle Gašpaříka a Koláře (2017) je ČR zapojena do těchto nákladních koridorů:

- koridor RFC 5. Gdyně – Katowice – Ostrava/Žilina – Bratislava/Vídeň/Klagenfurt – Udine – Benátky/Terst/Boloňa/Ravenna s větví Štýrský Hradec – Maribor – Lublaň – Koper/Terst,
- koridor RFC 7. **Praha** – Vídeň/Bratislava – Budapešť – Vidin – Sofie – Soluň – Atény s větví Bukurešť – Konstanca.
- koridor RFC 8. Bremerhaven/Rotterdam/Antverpy – Cáchy/Berlín – Varšava – Terespol (polsko-běloruská hranice) /Kaunas/Falkenberg – Praha/Vratislav – Katowice.
- koridor RFC 9. **Praha** – Horní Lideč – Žilina – Košice – Čierna nad Tisou (slovensko-ukrajinská hranice).

V rámci mezinárodní dopravy jsou důležitou funkcí pohraniční přechodové stanice. Těmito místy prochází železniční doprava do sousedících zemí. V rámci ČR jde o Rakouskou republiku, Slovenskou republiku, Polsko a Německou spolkovou republiku. Konkrétně jde podle Gašpaříka a Koláře (2017) o tyto hraniční přechody na železniční síti SŽ:

- **Rakouská republika:** Břeclav – Hohenau, Znojmo – Retz, České Velenice – Gmünd, Horní Dvořiště – Summerau.
- **Slovenská republika:** Mosty u Jablunkova – Čadca, Horní Lideč – Lúky pod Makytou, Lanžhot – Kúty, Vlárský průsmyk – Nemšová, Sudoměřice – Skalica, Hodonín – Holíč nad Moravou, Velká nad Veličkou – Vrbovce.
- **Polská republika:** Frýdlant v Čechách – Zawidów, Harrachov – Szklarska Poręba Górna, Královec – Lubawka, Meziměstí – Mieroszów, Lichkov – Międzyzlesie, Mikulovice – Głucholazy, Jindřichov ve Slezsku – Głucholazy, Bohumín – Chalupki, Petrovice u Karviné – Zebrzydowice, Český Těšín – Cieszyn.
- **Německá spolková republika:** Železná Ruda – Bayerisch Eisenstein, Česká Kubice – FurthimWald, Cheb – Schirnding, Aš – Selb– Plößberg, Vojtanov – Bad Brambach, Kraslice – Klingenthal, Potůčky – Johanngeorgenstadt, Vejprty – Bärenstein, Děčín – BadSchandau, Dolní Poustevna – Sebnitz, Rumburk – Ebersbach, Varnsdorf – Seifhennersdorf, Varnsdorf – Großschönau, Hrádek nad Nisou – Zittau.

3.4 Práce zabývající se zpracováním dat o infrastruktuře

Odborné práce, které by se zabývaly řešeným tématem této práce, není mnoho. Jde zejména o fakt, že data, jaká byla poskytnuta pro tuto práci, nebyla nikdy dříve zpracována. Výjimkou je pilotní článek využívající těchto dat od Dobešové a Kučery (2019), který jako jediný nese společné prvky s předkládanou diplomovou prací. Ostatní odborné a studentské práce využívají data podobné povahy anebo se zabývají tématy, ke kterým je možno tato data použít. Z odborných prací jde o Zprávu o zpracování Strategické hlukové mapy ČR od Šlachtové a kol. (2007). Ze studentských prací jde o Roubalíka (2017), který se ve své studentské bakalářské práci zabývá zpracováním dat z mýtných bran v ČR a Budilová (2019), která analyzuje nákladní železniční dopravu v oblasti Jesenicka.

3.4.1 Odborné práce

Dobešová a Kučera (2019) se ve svém článku zabývají zpracováním časových řad vytíženosti železniční infrastruktury ČR. Použity jsou obdobná data jako v této diplomové práci. Data jsou získávána za účelem sledování zatížení trati. Výsledky mají dva směry použití, prvním je podpora při rozhodování o údržbě infrastruktury a druhým je kalkulace ceny za pronájem železniční dopravní cesty. Článek si dává za cíl popsat v časovém horizontu chování v monitorovacích místech. Pro analýzu byly použity atributy počtu a váhy osobních a nákladních vlakových souprav v měsíčních úhrnech. Časový horizont analyzovaných dat byl od roku 2016 do roku 2018. Před analyzováním dat formou analýzy časové řady, byly data předzpracována. Konkrétně jde především o vyrovnání na ekvivalentní délku měsíce, tak aby data byla mezi sebou souměřitelná. Na řešení tohoto problému byly vyzkoušeny dva postupy. Prvním je přepočítání na rok s 360 dny s konstantní délkou měsíce 30 dní, druhým pak přepočítání na rok, který má

přesnou délku 366 dní. Pro vizuální analýzu vypracovaných grafu časových řad byla vybrána stanice Česká Třebová jako typicky reprezentant koridorového železničního uzlu. Dále pak stanice Břeclav jako typický představitel stanice poblíž frekventovaného železničního hraničního přechodu. Byly porovnány počty a váhy osobních a nákladních vlakových souprav v úhrnech za měsíce, kdy bylo zjištěno, že počet osobních vlakových souprav je mnohem větší než vlakových souprav nákladních. Naproti tomu váhy už netvořily tak velký rozdíl. Článek také uvádí, že existují výjimky, zejména na některých koridorových stanicích, místních trasách nebo hraničních přechodech. Byl zjištěn také velký rozptyl hodnot napříč měsíci a roky. V porovnání mezi druhy dopravy byly zjištěny větší sezónní výkyvy v nákladní dopravě, osobní doprava funguje v létě i v zimě podle platného jízdního řádu. Nákladní doprava závisí na objednávkách zákazníka. Trend v analýze časových řad byl zjišťován metodou klouzavých průměrů. Data grafy byly zpracovány v programu Microsoft Excel a možná predikce byla demonstrována v programu WEKA v. 3.8 s nadstavbou pro metodu Holt–Winters. Práce předpokládá další analýzy na těchto datech a uvažuje i součinnost s odborníky při další práci, zejména pak při zjišťování a odstranění vlivu mimořádností a výluk na výsledná data, která při možné zpracování jako časové řady vykazují výsledky s velkou volatilitou hodnot.

Šlachtová a kol. (2007) ve svém dokumentu *Zpráva o zpracování Strategické hlukové mapy ČR*, popisují využití dat o vytížení železniční infrastruktury pro zjištění hlukové zátěže ze železniční dopravy. Na základě *směrnice Evropského parlamentu a Rady 2002/49/ES* ze dne 25. června 2002 a následné související předpisy, která vymezuje základní principy, měla ČR povinnost do 30. června 2007 vypracovat hlukové mapy. Povinné jde dle této směrnice zmapovat aglomerace nad 250 000 obyvatel, hlavní silnice, po kterých projede více než 6 miliónů vozidel za rok, hlavní železniční tahy, po kterých projede více než 60 tisíc vlakových souprav za rok a pro hlavní letiště. Na základě § 80, odst. 1, písm. q) *Zákona č. 258/2000 Sb., o ochraně veřejného zdraví*, je pořízením strategické hlukové mapy pověřeno Ministerstvo zdravotnictví ČR. V první etapě zpracování byla vyhotovena strategická hluková mapa železnic. Za vyhotovení zodpovídala Národní referenční laboratoř pro užití GIS v ochraně a podpoře veřejného zdraví na Zdravotním ústavu se sídlem v Ostravě. Hlukové mapování bylo zpracováno pro 300 km železničních tratí splňujících kritéria pro železniční hlukové mapy uvedená výše. Na základě těchto pravidel byly vydefinovány jednotlivé ucelené úseky. Vstupní data o provozu železniční vozidel byla založena na aktuálním grafikonu. Tyto data spolu s daty o parametrech tratí byly získány od Výzkumného ústavu železničního, a.s. Dále byl získán digitální topografický model ČR od ČÚZK. Podrobnější informace o terénu v okolí železničních tratí byly získány ze zdrojů Střediska železniční geodézie Českých drah, a.s. Praha a Olomouc. Počty dotčených obyvatel v okolí byly získány od ČSÚ z databáze adresních bodů. Hluková mapa byla zpracována 1,5 km po obou stranách železniční tratě, bylo počítáno s hlukovým limitem daným *vyhláškou č.523/2006 Sb. O hlukovém mapování, §2, odst. 3.*, která definuje pro železniční dopravu mezní hodnotu hlukového indikátoru pro den–večer–noc L_{dvn} rovnou 70 dB a pro noc L_n rovnou 65 dB. Pro výpočet hluku je ve směrnici 2002/49/ES doporučena Nizozemská metoda, která je v českém překladu dostupná na stránkách Ministerstva dopravy, jako *Pokyny k výpočtu a měření hluku ze železniční dopravy z roku 1996*. Mapové výstupy ze strategického hlukového mapování byly pro železnice zpracovány pro varianty L_{dvn} a L_n ve formátu A0 v měřítku 1:30 000 a ve formátu A2 v měřítku 1:10 000.

3.4.2 Studentské práce

První zmíněnou studentskou prací zabývající se zpracováním dat o infrastruktuře je Roubalík (2017). Tato práce je založena na zpracování dat z mýtných bran spravovaných v ČR firmou KAPSCH. Data vznikají na bázi mikrovlnného snímání pohybu vozidel na komunikacích. Všechna dotčená vozidla (nad 3,5 t) musí být vybavena palubní jednotkou pro komunikaci s mýtnými branami, která je nositelem parametrů a identifikace vozidla. Na základě parametrů je vyměřována konkrétní výše mýta a všechny údaje jsou ukládány do databáze. Celý systém funguje pod záštitou zmiňované firmy KAPSCH a Ředitelství silnic a dálnic (ŘSD). Od této instituce byla tato data nakonec získána pro analýzy na katedře geoinformatiky Univerzity Palackého v Olomouci. Práce si pokládá za cíl zjistit proměnlivost dopravního toku v řádu dne i týdne v závislosti na různém období v rámci roku. V první části se práce věnuje analýze dopravních toků napříč všemi mýtnými branami v ČR. Cílem bylo zjistit maximální a minimální hodnoty zatížení celých komunikací i dílčích úseků. Tyto informace napomohli ke zjištění času dopravní špičky a dopravního sedla. Protože data obsahují také informace o autobusových dopravcích, tak se práce v druhé části věnuje ovlivněním celkového zatížení dopravy jednotlivými autobusovými dopravci.

Budilová (2019) se ve své bakalářské práci zabývá analýzou nákladní železniční dopravy v oblasti Jesenicka. Tato práce se soustředí na popis současné situace transportu napadeného dřeva kůrovcem v oblasti Jesenicka s ohledem na geografické poměry území. V další části poté představuje nový návrh řešení této problematiky a srovnává jej s analýzou provedenou na ČVUT v roce 2015.

3.5 Shlukování

Shlukováním, jako jednou z metod dataminingu, se zabývá Šarmanová (2012) ve své publikaci *Metody analýzy dat*. Publikace se nejdříve věnuje popisu rozdílů mezi klasifikací a shlukováním. Lze říci, že u klasifikace známe předem kritérium, podle které se má množina rozřadit do kategorií. U shlukování jde o zjištění těchto kritérií, která nejsou předem známá. Na základě volby vhodné shlukovací metody poté dostáváme hledané klasifikační třídy. Následně publikace definuje shlukovací metody, úlohy, stručnou historii shlukování jako metody a problémy, které mohou nastat při nevhodném přístupu ke shlukovací analýze. Dále řeší výběr relevantních atributů a podobností objektů. Představuje koeficienty podobnosti a vzdálenosti. Ve dvou podkapitolách potom rozděluje shlukování na hierarchické a nehierarchické. Nehierarchické hledá nejlepší prostý rozklad množiny objektů na shluky, tedy na podmnožiny. Shlukování hierarchické nehledají pouze rozklad objektů na shluky, ale dávají jako výsledek i hierarchii takovýchto shluků. Kapitola shluková analýza si dává za cíl popsat problémy shlukovací analýzy, popsat typy shluků, popsat typy shlukovacích metod a jimi řešitelných úloh, zhodnocení vhodnosti úlohy pro shlukování a na závěr uvádí konkrétními případy provedení shlukové analýzy. Dalším, který se věnuje problémem shlukování v dataminingu je Klímek (2008), který publikoval článek *Shlukovací metody v dataminingu* ve vědeckém časopise *E+M Ekonomie a Management*. Autor se v článku nejdříve zabývá obecnou definicí dataminingu, kde přibližuje různé směry, ke kterým je možno dataminingu použít. Jedním z těchto směrů je i shluková analýza, které se věnuje ve zbytku článku. Nejdříve se věnuje definici shlukovací analýzy, poté mírou vzdálenosti mezi objekty, představením shlukovacích metod a na závěr praktickými příklady použití shlukové analýzy.

3.6 Časové řady

Jednou z nejkompexnějších publikací na téma časové řady je *Analýza časových řad* (Křivý, 2012). V úvodu publikace se zabývá definicí pojmu časové řady, vysvětlením souvisejících pojmů, základními přístupy při analýze časových řad a problémy, které mohou při řešení nastat. Dále rozebírá téma více do hloubky. Řeší rozklad časových řad na složky různými metodami a postupy. Následně vysvětluje lineární modely stacionárních, nestacionárních, sezónních a vícerozměrných časových řad. Na závěr se zaměřuje na základní pojmy spektrální analýzy časových řad. Další publikací zabývající se výhradně tématem časových řad je *Úvod do analýzy časových řad* od Hančlové a Tvrdého (2003). V první části publikace se autoři soustředí na vysvětlení základních pojmů v analýze časových řad, jako jsou druhy časových řad, grafická analýza a popisné charakteristiky. Dále se v této části věnují základní úpravě časových řad, jež obsahuje doplnění chybějících hodnot, transformaci měřítka kombinace časových řad, časový posun, sezónní difference, kumulativní součet a vyhlazování časových řad. Na závěr první části se autoři zaměřují na problémy časových řad a metody analýzy časových řad. Ve druhé části je předchozí teorie vysvětlena případovou studií *Analýza míry nezaměstnanosti v okrese Karviná*. Druhým příkladem je poté vysvětlení postupu analýzy ukazatelů na úrovni obce. Litschmanová (2010) se ve své publikaci *Úvod do analýzy časových řad* věnuje základním pojmům z oblasti časových řad a přináší vysvětlení všech těchto pojmů formou mnoha příkladů. Publikace je rozdělena na jednotlivé kroky analýzy časových řad a ke každému kroku přidá jeden nebo více příkladů důležitých pro jeho pochopení. K příkladům je dostupné také řešení. Klíček a Kasal (2007) se ve své publikaci *Počítačové zpracování dat v programu Statistica* ve třetí kapitole věnují problematice zpracování časových řad. Zejména je zde dobře vysvětlena samotná definice časových řad, jejich rozklad na jednotlivé složky a vyhledání trendu. Tuto publikaci je možné využít jako obecný zdroj k problematice časových řad, i když pro řešení této problematiky používá konkrétní software.

4 PŘEDZPRACOVÁNÍ DAT

Před analýzou poskytnutých dat, bylo nutné tyto data předpřipravit, aby bylo možné s nimi dále pracovat. Oba poskytnuté datasety byly v nevhodné formě pro analýzy. Data pro monitorovací body neměli GPS souřadnice a k jednotlivým monitorovacím bodům bylo více záznamů s číselnými hodnotami. To znamená, že se v tabulce opakovali záznamy se stejným názvem monitorovací bodu, ale pokaždé s jiným měsíčním úhrnem. Šlo o částečné součty různých druhů dopravy během daného měsíce, proto se konkrétní druhy dopravy v záznamech s hodnotami pro jeden měsíc často opakovaly. Úseková data, na rozdíl od bodových, mají polohovou informaci, bohužel uloženou ve špatném formátu, který bylo nutné pro tvorbu prostorové vrstvy upravit. Velké množství těchto polohových informací bylo chybně zapsáno, nebo zde byly překlepy, a tudíž bylo nutné tyto monitorovací body upravit tak, aby měli správnou polohu. Dále bylo nutné vytvořit primární klíč pro připojování dat k jednotlivým úsekům. Číselná data obou datasetů bylo nutné pro získání měsíčních a ročních úhrnů transformovat pomocí kontingenční tabulky.

4.1 Data pro jednotlivé monitorovací body

Tyto data jsem použil pro kategorizaci železniční stanic za celé území ČR, kategorizováno bylo na základě kombinace atributů měsíčních a ročních úhrnů v roce 2016. Rok 2016 byl vybrán, protože jde o kompletní časovou řadu po měsících za celý rok. Z tohoto důvodu nejsou použitelná data za rok 2017, protože chybí poslední 2 měsíce a vzhledem k tomu není součet těchto dat souměřitelný s rokem 2016. Vzhledem k následné vizualizaci bylo nutné získat polohovou složku těchto dat, která v nich nebyla obsažena. Monitorovací body jsou totožné s daty za úseky, a tudíž z nich byly odvozeny. Při tvorbě prostorové vrstvy byl zvolen následující postup. Nejdříve byla vytvořena tabulka se souřadnicemi, názvem monitorovacího bodu a primárním klíčem, který je definován jako kód s interním označením SŽ dle číselníků SR70. Poté byly připraveny další 2 typy tabulek pro spojení s první tabulkou, ve kterých byly agregovány číselné hodnoty za rok a měsíce. První tabulka se poté nahrála do softwaru ArcGIS Pro a vytvořila se z ní prostorová vrstva ze souřadnic a poté k ní byly připojovány tabulky s číselnými hodnotami.

První tabulka byla vytvořena vykopírováním a úpravou hodnot z tabulky s úsekovými daty pro rok 2016. Do nové tabulky v Excelu byly do třech sloupců pod sebe umístěny pole počátečních a koncových stanic. Konkrétně šlo o název monitorovacího bodu, kód SR70 a souřadnice. Tento postup byl nutný, protože v úsekových datech jsou některé monitorovací body obsažené pouze v počátečních a některé pouze v koncových atributech, třetím případem je možnost, kdy jsou monitorovací body umístěné v obou. Druhým důvodem je ten, že každý monitorovací bod má v tabulce hned několik záznamů. Vzhledem k tomu že bylo potřeba získat pouze popisné informace a polohu monitorovacího bodu bylo možné jednoduše smazat duplicitní řádky ve výsledné tabulce a získat seznam všech monitorovacích bodů a jejich souřadnic v neupraveném tvaru. Poté bylo potřeba provést úpravu souřadnic, které jsou ve tvaru $ESS^{\circ}MM'VV'' - NSS^{\circ}MM'VV''$ v jednom sloupci. Pro využití funkce *display XY data* v aplikaci *ArcGIS Pro* bylo nutné data o poloze transformovat do desítkové soustavy ve dvou sloupcích pro souřadnice x a y. Byly vymazány mezery v původním sloupci GPS souřadnice, a poté nahrazeny znaky stupňů, minut a vteřin mezerami. Tyto mezery byly

použity jako dělicí znak, při dělení řádů do jednotlivých sloupců, pomocí funkce text do sloupců, v programu Excel. Následně byl proveden výpočet, kdy byly pro osy x a y sečteny stupně s minutami dělenými 60 a vteřinami dělenými 3600. Vznikla tak vhodná struktura tabulky pro tvorbu prostorové vrstvy v ArcGIS Pro, ze souřadnic obsažených v tabulce. Po vložení tabulky do ArcGIS Pro, byla aplikována metoda *Display XY data* a výsledná vrstva byla uložena jako nová *feature class*. Následně byly odstraněny monitorovací body s chybějícími souřadnicemi nebo souřadnicemi mimo hranice ČR. Data mimo hranice ČR byla odstraněna na základě prvního dvojčíslí v kódu SR70. Pokud bylo dvojčíslí 54, tak šlo o data na území ČR, pokud bylo jiné, byla data z území okolních států. Tabulka s hodnotami pro rok a měsíce byla získána v programu Excel pomocí kontingenční tabulky (viz obr. 4.1). Pro roční úhrny hodnot všech číselných atributů byla vytvořena kontingenční tabulka, ve které byl umístěn do volby *řádky*, sloupec kódu SR70, jako primárního klíč pro následné spojení s prostorovou vrstvou Do volby *sloupce*, byl umístěn řádek s názvy atributů s číselnými hodnotami. Do volby *hodnoty*, byl umístěn výběr všech číselných hodnot. Dále pro možnost filtrování dat na nákladní a osobní dopravu byla přidána do volby *filtry* možnost filtrovat dle atributu příznak nákladní vlakové soupravy. Možnost 0 znamená vlak osobní a možnost 1 vlak nákladní. V možnosti pro souhrn dat ve volbě hodnoty byl zvolen součet pro absolutní čísla a průměr pro průměrné hodnoty.

Popisky řádků	Sum of počet_vlaku	Sum of hmotnost_soupravy	Sum of počet_vozu_soupravy	Average of prumerna_delka_soupravy	Average of prumerny_pocet_naprav
51-052894-01	10005	749697	11048	37,55555556	5,444444444
51-057109-00	3512	651891	16040	132,5045455	32,43272727
51-057109-01	3512	651891	16040	132,5090909	32,43636364
51-057109-02	3512	651891	16040	132,5090909	32,43636364
51-057117-00	3512	651891	16040	132,5090909	32,43636364
51-067900-00	1	929	32	430	128
51-075507-00	20	10809	242	267	52,86111111
51-075580-00	1	85	1	17	4
51-075648-00	2	252	3	24,5	6
51-075655-01	2	252	3	24,5	6
51-075655-02	2	252	3	24,5	6
51-075705-00	5	2082	36	197,6666667	31
51-075796-00	11	5714	108	279,7142857	45,85714286
51-075903-00	5	2082	36	197,6666667	31
54-007013-00	9536	442061	18828	21,1627907	6,139534884
54-300004-00	1114	1556993	26233	218,034111	61,59010781
54-300012-00	1190	1409880	29678	238,9407913	70,26321985
54-300020-00	2716	3171629	64148	221,616895	63,51004566
54-300038-00	7834	9301743	172678	200,898676	60,25670833
54-300046-00	2983	786959	16061	63,41038961	18,22662338
54-300053-00	2222	1463099	30229	270,8151096	66,07741201
54-300061-00	3297	4656950	82924	379,5078706	98,25836898
54-300079-00	4344	2956100	49106	250,4428312	55,61161529

Obr. 4.1 Náhled kontingenční tabulky ročních úhrnů pro data za jednotlivé monitorovací body v programu Excel.

V případě tabulky pro data za měsíce je postup obdobný. Pro agregaci z původní tabulky byla použita kontingenční tabulka (viz obr. 4.2). Do volby *řádky* byl opět zvolen sloupec s kódem SR70, do volby sloupce řádek s označením období ve formátu RRRRMMDD a do volby *hodnoty* byly zvoleny sloupce všech číselné hodnoty. Do volby *filtry* byl opět použit příznak nákladní vlakové soupravy a byly zvoleny stejné možnosti pro souhrn dat, tedy pro průměrné hodnoty průměry a pro absolutní hodnoty součty.

Popisky řádků	20160101	20160301	20160401	20160501	20160601	20160701	20160801	20160901	20161001	20161101	20161201 (prázdná)	Celkový součet	
51-052894-01	847	796	852	822	850	822	848	852	824	849	824	819	10005
51-057109-00	293	281	292	291	293	301	297	320	307	300	260	277	3512
51-057109-01	293	281	292	291	293	301	297	320	307	300	260	277	3512
51-057109-02	293	281	292	291	293	301	297	320	307	300	260	277	3512
51-057117-00	293	281	292	291	293	301	297	320	307	300	260	277	3512
51-067900-00		1											1
51-075907-00		1	3			1		7	4	3	1		20
51-075580-00								1					1
51-075648-00								1		1			2
51-075655-01								1		1			2
51-075655-02								1		1			2
51-075705-00								1	3	1			5
51-075796-00								3	4	3	1		11
51-075903-00									1	3	1		5
54-007013-00	722	713	743	636	775	843	1061	1108	802	716	693	724	9536
54-300004-00	92	86	112	94	93	75	80	102	61	124	94	101	1114
54-300012-00	129	55		66	88	93	99	104	182	159	136	79	1190
54-300020-00	106	191	204	305	261	223	236	206	243	230	234	187	2716
54-300038-00	736	770	640	531	663	625	606	619	642	706	748	548	7834
54-300046-00	240	252	282	238	246	263	238	253	237	256	241	237	2983
54-300053-00	172	198	202	206	197	164	189	197	186	192	187	132	2222
54-300061-00	263	254	280	300	330	260	285	265	293	237	278	252	3297

Obr. 4.2 Náhled kontingenční tabulky měsíčních úhrnů pro data za jednotlivé monitorovací body v programu Excel.

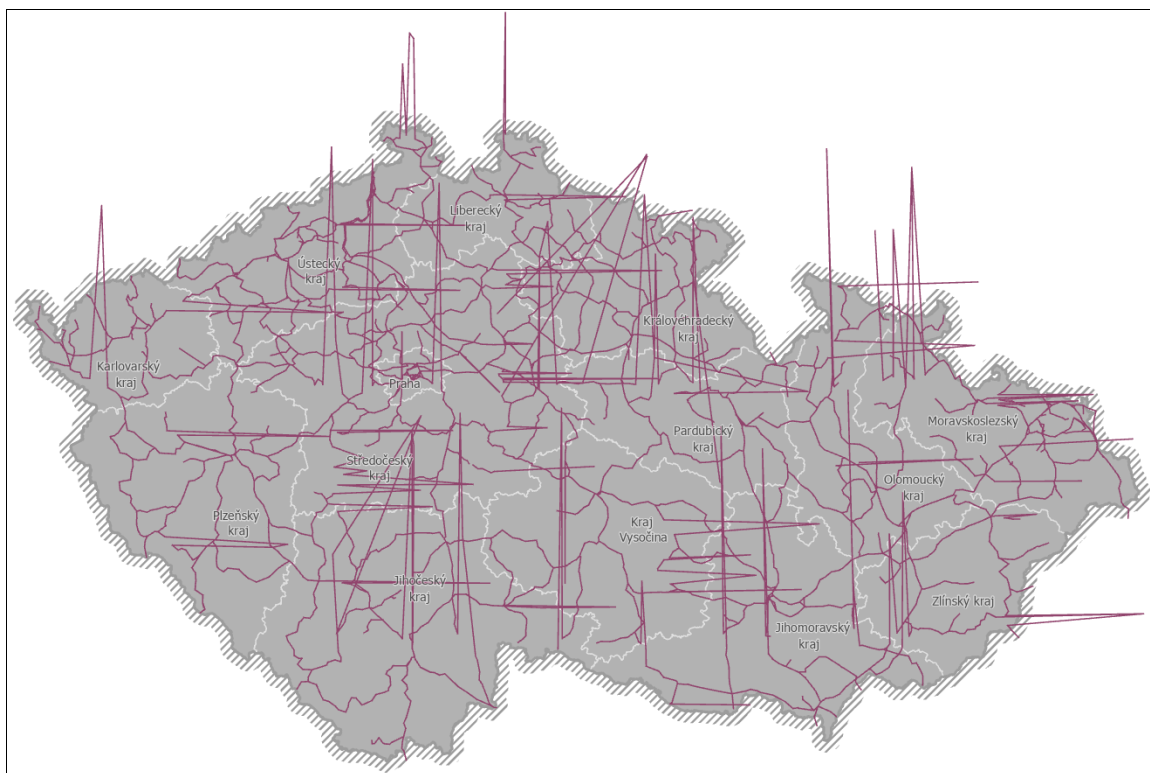
Pro připojení hodnotové tabulky k *feature class* v programu ArcGIS Pro byla vybrána potřebná agregace, konfigurace dopravy a výběr atributů. Poté byla vytvořena tabulka se všemi náležitosti pro připojení. Poté už byla připojena výsledná tabulka a uložena *feature class* s připojenými daty jako novou *feature class*, která již tato data obsahuje.

4.2 Úseková data

Tyto data byla použita primárně na analýzu časových řad vybraných typů stanic na základě prvotních vizuálních analýz dat a také pro tvorbu zjednodušené železniční sítě a její kategorizaci na základě číselných atributů v úhrnech za rok. Použity byla kompletní data, protože jde o kompletní agregace za měsíce od roku 2016 do roku 2019. Data byla opět nutné transformovat, protože obsahovala několik záznamů pro jeden monitorovací bod. Opět bylo, jako u předchozích dat, použito zpracování na prostorovou tabulku a tabulku s číselnými daty. Zpracována byla verze pro měsíční agregace jako vstup do časové analýzy a roční agregace jako vstup do tvorby a kategorizace zjednodušené železniční sítě. Úprava datových tabulek byla totožná s postupem z předchozího zpracování dat pro jednotlivé monitorovací body. Struktura prostorové tabulky byla rozdělena dle struktury vstupních dat na úseky, přičemž byly informace o poloze zdvojené, vždy pro počáteční a koncovou stanici. Navíc byl vytvořen primární klíč z kombinace kódů SR70.

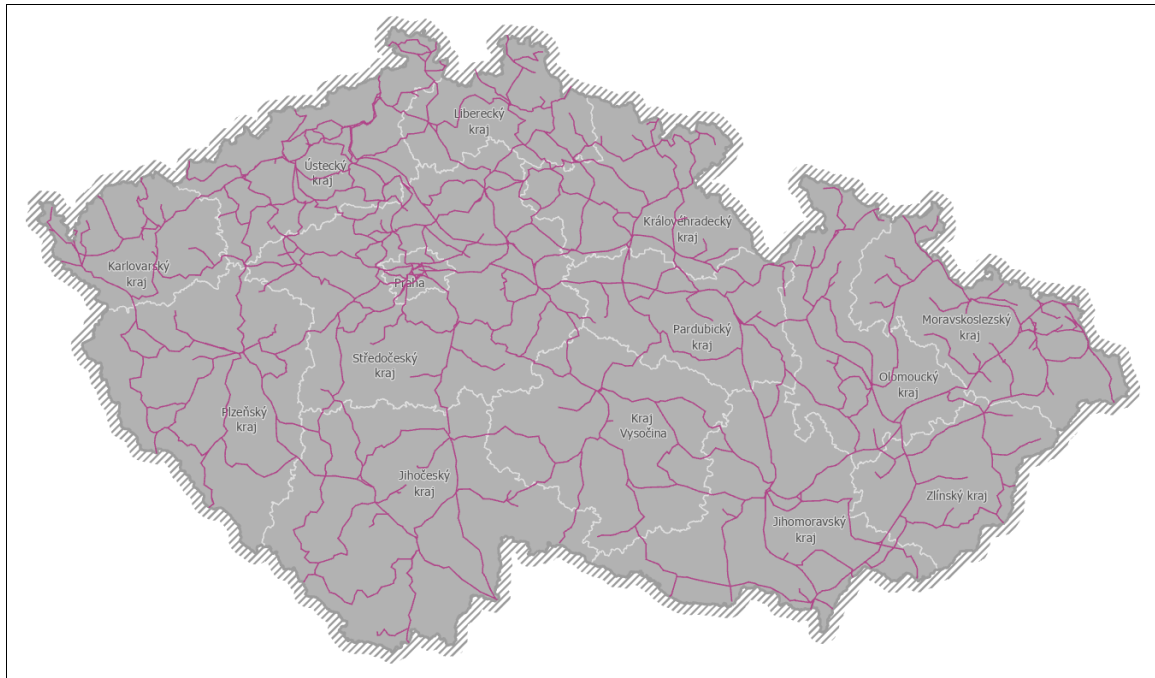
Nejdříve byla vytvořena tabulka, která sloužila jako vstup pro tvorbu zjednodušeného síťového modelu. Síťový model byl vytvořen z dat pro rok 2019. Nejdříve byly do nové tabulky vykopírovány kódy SR70 počátečních a koncových stanic a jejich souřadnice. Byly odstraněny duplicity vzniklé několika záznamy pro jednotlivé monitorovací body za jednotlivé měsíce. Primární klíč byl vytvořen spojením kódů stanic do jednoho sloupce a rozdělen podtržítkem. Souřadnice byly upraveny stejným způsobem jako u dat za jednotlivé monitorovací body. Když byly souřadnice ve formátu vhodném pro vytvoření prostorové vrstvy, tak byla tabulka otevřena v programu ArcGIS Pro. Dále byl použit nástroj *XY to Line*, který využil struktury dat na počáteční a koncové monitorovací body a vznikl liniový schematický model, železniční sítě dle úsekových dat za rok 2019. Souřadnice v původní tabulce byly očištěny od takových monitorovacích bodů, které měli nulovou hodnotu. Po vytvoření modelu bylo zřetelné, že některé monitorovací body jsou umístěny zcela mimo svojí správnou polohu (viz obr. 4.3). Tento problém byl způsoben zejména vypouštěním nuly na pozici minut v zápisu úhlových souřadnic. Po

přepočítání na desetinnou úhlovou míru zde vznikaly chyby. V méně případech pak polohová informace zcela chyběla.



Obr. 4.3 Zjednodušený liniový model před úpravou souřadnic v programu ArcGIS Pro.

Tyto monitorovací body byly vyhledány a byly opraveny jejich souřadnice v původní vstupní tabulce. Poté byl znovu vytvořen liniový model (viz obr. 4.4). Dále bylo nutné odstranit monitorovací body mimo území ČR, toho bylo docíleno obdobně jako u dat pro jednotlivé monitorovací body výběrem dat, které mají v kódu sr70 první dvojčíslí 54. Takto hotový liniový model (viz obr. 4.4) byl připravený k připojení tabulky s číselnými atributy agregovanými za rok i měsíc.



Obr. 4.4 Zjednodušený liniový model po úpravě souřadnic v programu ArcGIS Pro.

Pro připojení číselných dat k prostorové vrstvě byly vytvořeny, stejně jako pro data za jednotlivé monitorovací body, 2 typy tabulek. První seskupuje data v agregacích za rok a druhá v agregacích za měsíce. Jediným rozdílem oproti datům za monitorovací body je ten, že tato data jsou vztažena k jednotlivým úsekům mezi nimi. Data byla získána, opět jako u dat za monitorovací body, za pomoci vhodného použití kontingenčních tabulek v Microsoft Excel.

Data, která sloužila jako vstup do analýzy časových řad, byla získána také pomocí kontingenční tabulky (viz obr. 4.5). Byly použita data za všechny roky, tedy datové sady pro rok 2016 až 2019 včetně. Jejich úprava probíhala stejným způsobem. Pomocí kontingenční tabulky byly přepočítány data za jednotlivé měsíce pro jednotlivé úseky za roky. Data je možné filtrovat dle druhu dopravy na osobní, nákladní, soupravný, lokomotivní a služební.

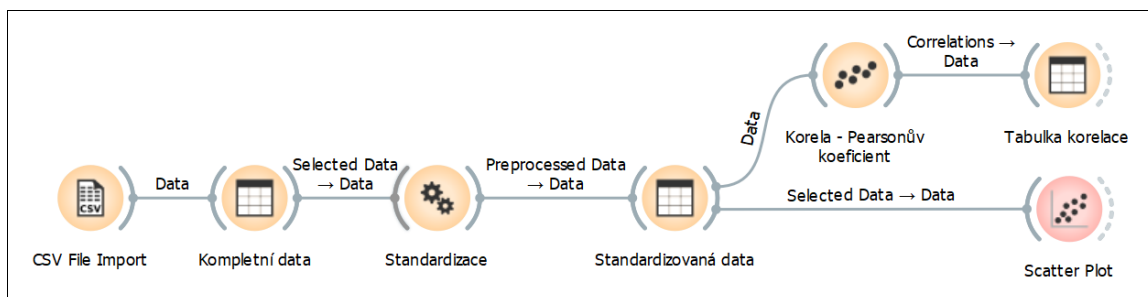
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Kategorie dopravy	(vše)													
2															
3	Součet z Počet vlaků														
4	Popisky sloupců	20190101	20190201	20190301	20190401	20190501	20190601	20190701	20190801	20190901	20191001	20191101	20191201 (prázdné)	Celkový	
5	51-052704-00_54-506725-00	47	44	57	39	24	21	26	19	33	41	35	22		
6	51-052894-01_54-505321-00	972	880	974	944	978	944	976	977	940	972	952	442		
7	51-052894-01_80-098707-00	817	740	819	793	823	793	820	822	790	814	797	370		
8	51-055400-00_54-507202-00				8	44	43	36	40	36					
9	51-055509-00_54-502906-00	216	196	217	204	178	177	190	187	208	217	208	215		
10	51-056226-00_54-509216-00	225	240	217	182	206	227	297	299	175	178	142	168		
11	51-057109-00_51-057109-02	291	270	300	291	122	35								
12	51-057109-00_51-057117-00	291	270	300	291	122	214	294	292	282	284	282	275		
13	51-057109-00_51-637520-00						179	294	292	282	284	282	275		
14	51-057109-01_51-057109-02	291	270	300	291	122	35								
15	51-057109-01_54-300426-00	291	270	300	291	122	35								
16	51-057117-00_54-301549-00	291	270	300	291	122	214	294	292	282	284	282	275		
17	51-057208-00_54-504407-00	191	172	187	173	181	179	143	180	168	188	194	183		
18	51-067900-00_54-300244-00	391	363	351	305	371	416	443	436	376	403	377	464		
19	51-067900-00_54-300251-00	595	581	683	605	641	621	598	571	605	568	588	637		
20	51-073973-00_54-505321-00												544		
21	51-073973-00_80-215046-00												431		
22	51-075507-00_51-07580-00												1		
23	51-075507-00_54-300343-00	797	773	889	913	926	833	841	880	845	889	846	750		
24	51-075580-00_51-075648-00												1		
25	51-075648-00_51-502740-00												1		
26	51-075655-01_54-300145-00	380	382												
27	51-502732-00_51-502740-00												1		
28	51-502732-00_54-300145-00			41	446	431	421	370	373	281	270	360	342	353	

Obr. 4.5 Náhled kontingenční tabulky měsíčních úhrnů pro data za úseky v programu Excel.

4.3 Výběr atributů

Data za jednotlivé monitorovací body i data za úseky obsahují stejné číselné atributy popisující zatížení železniční infrastruktury. Ne všechny atributy jsou však vhodné pro další zpracování základních vizualizací nebo analýz. Pro vizualizace je vhodné využít takových atributů, které dobře popisují situaci a nekorelují spolu, protože v tom případě by vznikaly výstupy s podobnou informační hodnotou. Pro kategorizaci je nutné provést rozhodnutí, které atributy popisují zatížení infrastruktury a které jsou pouze dalšími doplňujícími prvky, které už z tohoto hlediska nepřinášejí žádné další informace. Vzhledem k povaze analýzy časových řad, kdy vybíráme jednu informaci, která bude popisovat chování jednoho prvku za určitý čas, potřebujeme takový atribut, který bude nejlépe vystihovat zatížení železniční infrastruktury.

Atributy pro prvotní vizualizace byly vybrány na základě míry podobnosti jednotlivých atributů. V programu Orange byl vytvořen model pro výpočet korelace všech atributů (viz obr. 4.6) a korelace váhy a hmotnosti osobních a nákladních vlakových souprav. Do programu Orange byla nahrány data ve formátu CSV, poté byla standardizována a následně byl proveden výpočet korelace na základě Pearsonova koeficientu, který slouží pro výpočet korelace reálných dat. Pro názornost byl použit korelační diagram, též známý jako *Scatter plot*.



Obr. 4.6 Model pro výpočet korelace v programu Orange.

Výsledné hodnoty kombinací všech vstupních atributů bez rozlišení dopravy, byly výsledné korelace následující (viz obr. 4.7).

1	+1.000	Součet z Hmotnost souprav	Součet z Počet vlaků
2	+0.990	Průměr z Průměrný počet náprav	Průměr z Průměrný počet vozů
3	+0.965	Průměr z Průměrná hmotnost soupravy	Průměr z Průměrný počet náprav
4	+0.965	Průměr z Průměrná délka soupravy	Průměr z Průměrný počet náprav
5	+0.960	Průměr z Průměrná délka soupravy	Průměr z Průměrný počet vozů
6	+0.953	Průměr z Průměrná hmotnost soupravy	Průměr z Průměrný počet vozů
7	+0.925	Průměr z Průměrná délka soupravy	Průměr z Průměrná hmotnost soupravy
8	+0.024	Průměr z Průměrná délka soupravy	Součet z Hmotnost souprav
9	+0.023	Průměr z Průměrná hmotnost soupravy	Součet z Hmotnost souprav
10	+0.023	Průměr z Průměrný počet náprav	Součet z Hmotnost souprav
11	+0.023	Průměr z Průměrná délka soupravy	Součet z Počet vlaků
12	+0.022	Průměr z Průměrný počet vozů	Součet z Hmotnost souprav
13	+0.021	Průměr z Průměrná hmotnost soupravy	Součet z Počet vlaků
14	+0.021	Průměr z Průměrný počet náprav	Součet z Počet vlaků
15	+0.020	Průměr z Průměrný počet vozů	Součet z Počet vlaků

Obr. 4.7 Vypočtené korelace všech kombinací vstupních atributů bez rozlišení dopravy v programu Orange.

Největší míru korelace zaznamenali očekávaně atributy celkového počtu a hmotnosti vlakových souprav. Zde je nutné podotknout, že výpočet korelace proběhl bez rozlišení typu dopravy. Tyto atributy jsou nositeli zajímavých informací, zejména pokud porovnáváme nákladní a osobní dopravu. Pokud jde o porovnání absolutních atributů mezi sebou a relativních mezi sebou, je zde velmi vysoká míra korelace. Pokud porovnáváme atributy relativní s absolutními je korelace velmi malá. Vzhledem k tomu, že pro zatížení železniční infrastruktury jsou důležité zejména váha a počet vlakových souprav, byly pro další srovnání s rozlišením typu dopravy vybrány právě tyto atributy. Níže (obr. 4.8) je možné vidět, že pokud je typ dopravy stejný, je očekávatelné, že korelace bude velmi vysoká. Zatímco při porovnání napříč typy dopravy se tato korelace snižuje. Obecně platí, že osobních vlakových souprav je násobně větší množství než nákladních, ale nákladní mají na jednu vlakovou soupravu mnohem vyšší hmotnost než vlakové soupravy osobní.

1	+0.992	hmotnost_vlaku_nakl	pocet_vlaku_nakl
2	+0.958	hmotnost_vlaku_osob	pocet_vlaku_osob
3	+0.675	hmotnost_vlaku_osob	pocet_vlaku_nakl
4	+0.627	hmotnost_vlaku_nakl	hmotnost_vlaku_osob
5	+0.615	pocet_vlaku_nakl	pocet_vlaku_osob
6	+0.568	hmotnost_vlaku_nakl	pocet_vlaku_osob

Obr. 4.8 Vypočtené korelace počtu a hmotností osobních a nákladních vlakových souprav v programu Orange.

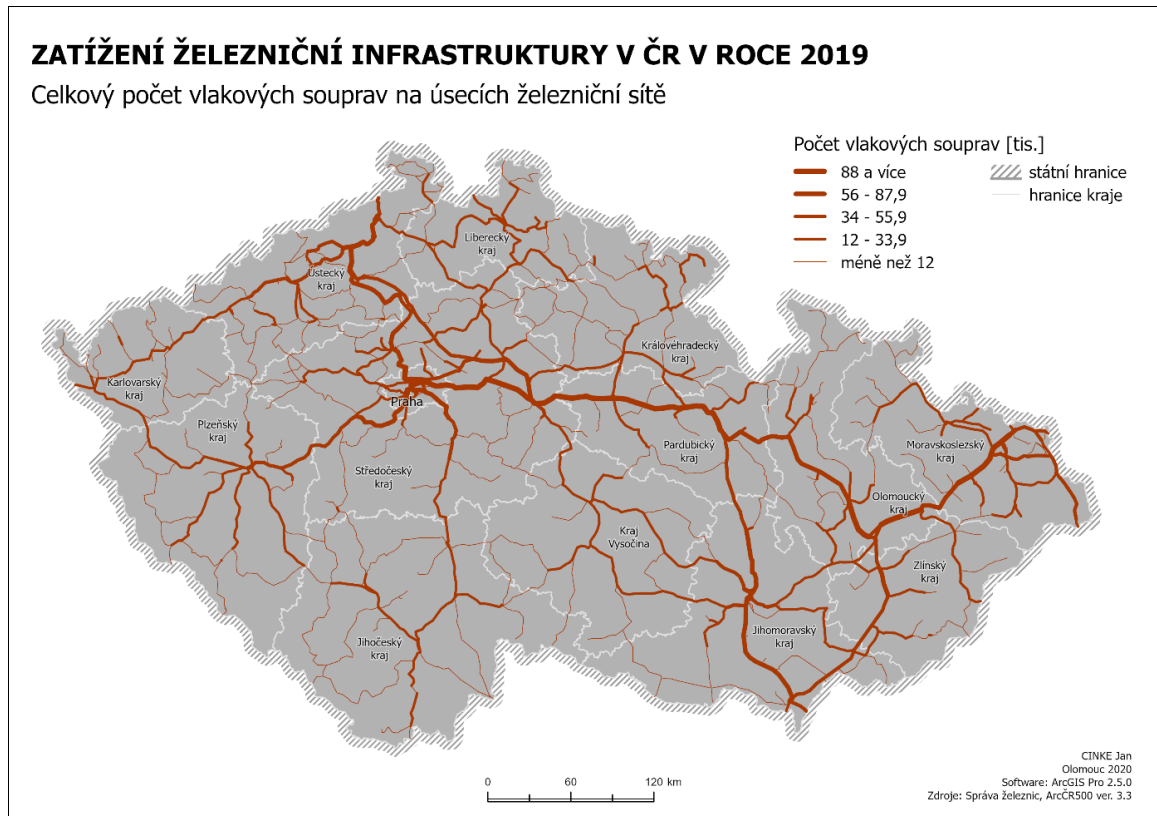
V případě výběru atributů pro kategorizaci železničních úseků a stanic bude použito stejného faktu jako u výběru atributů pro prvotní analýzy. V rámci shlukování budou použity kombinace atributů celková váha a počet osobních a nákladních vlakových souprav. Nebudou zde uvažovány relativní atributy, které do zatížení nepřinášejí příliš mnoho dalších informací a některé jsou pouze odvozením od absolutních atributů, například průměrná hmotnost.

5 ZÁKLADNÍ VIZUALIZACE

Vzhledem k tomu, že byly z poskytnutých dat vytvořeny prostorové vrstvy, bylo vhodné před další analýzou vytvořit základní vizualizace pro lepší pochopení vlastností dat, které nejsou zřetelné z pouhého zápisu v tabulce. Vizualizace pomohly k lepšímu pochopení studovaných jevů a jejich názornosti, při hledání souvislostí, které mohou být důležité při rozhodování o dalším zpracování dat. Nejdříve byly vytvořeny vizualizace pomocí jednoduchých mapových výstupů obsahujících počty a hmotnosti osobních a nákladních vlakových souprav. Další tematický okruh bylo procentuální vyjádření převládajícího typu dopravy z hlediska váhy a počtu vlakových souprav. Pro názornost měsíčního průběhu hodnot nákladní a osobní dopravy, byly vypracovány grafy ročního průběhu hodnot hmotnosti a počtu vlakových souprav, pro vybrané monitorovací body.

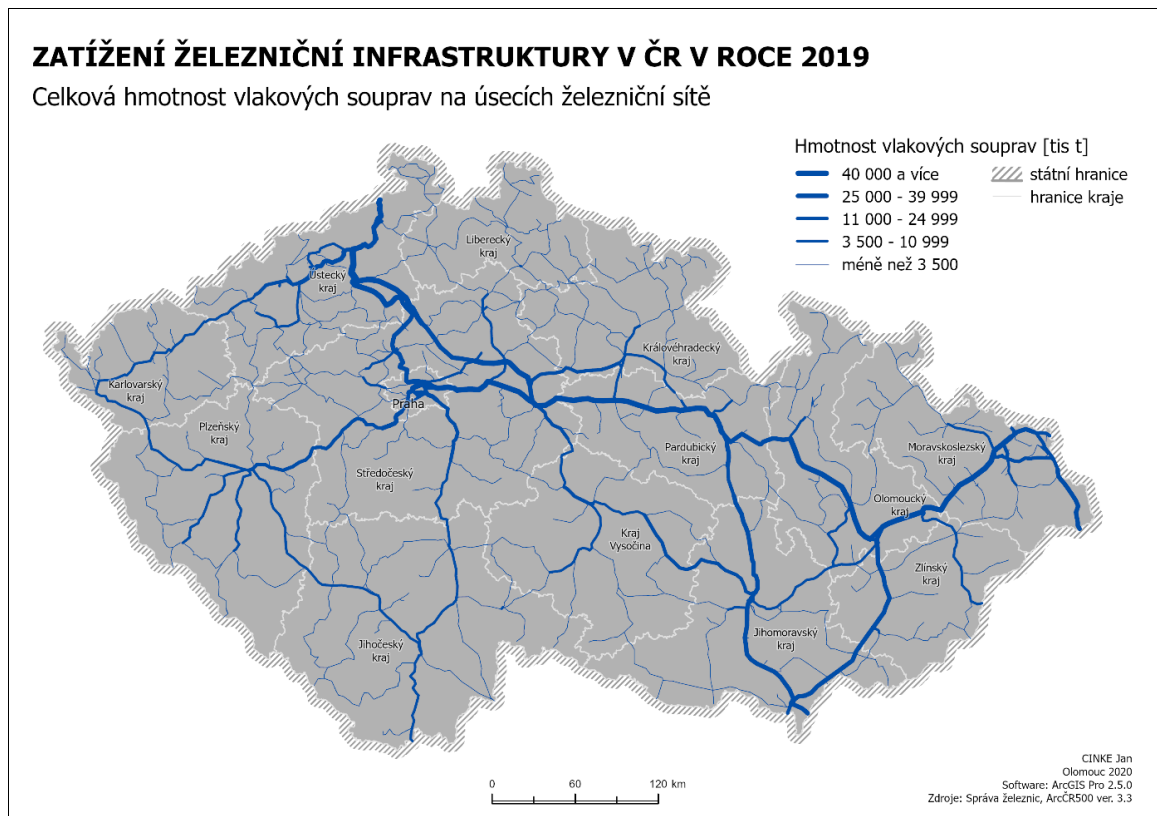
5.1 Mapy

V první části byly vytvořeny mapové výstupy pro celkovou hmotnost a počet nákladních vlakových souprav za rok 2019 v ČR. Tyto výstupy poskytují základní seznámení se zatížením železniční infrastruktury v ČR na všech tratích vlastněných Správou železnic reprezentovaných ve výstupech zjednodušeným schematickým modelem, který byl vytvořen na základě spojení jednotlivých monitorovacích míst. Na první mapě (viz obr. 5.1) je zobrazen počet vlakových souprav, rozdělený do 5 kategorií. Je zde vidět očekávatelný fenomén, kdy největší počty vlakových souprav proudí po tranzitních koridorech. Zejména jde o 1. Tranzitní koridor v jeho plné délce z Děčína přes Prahu, Českou Třebovou, Brno a Břeclav. Dále pak 2. Tranzitní koridor z Petrovic u Karviné přes Přerov do Břeclavi. 3. Tranzitní koridor je poté zatížen velkým počtem vlakových souprav zejména v úseku z Prahy do Ostravy přes Olomouc. V případě 4. Tranzitního koridoru jde o úsek mezi Děčínem a Prahou.



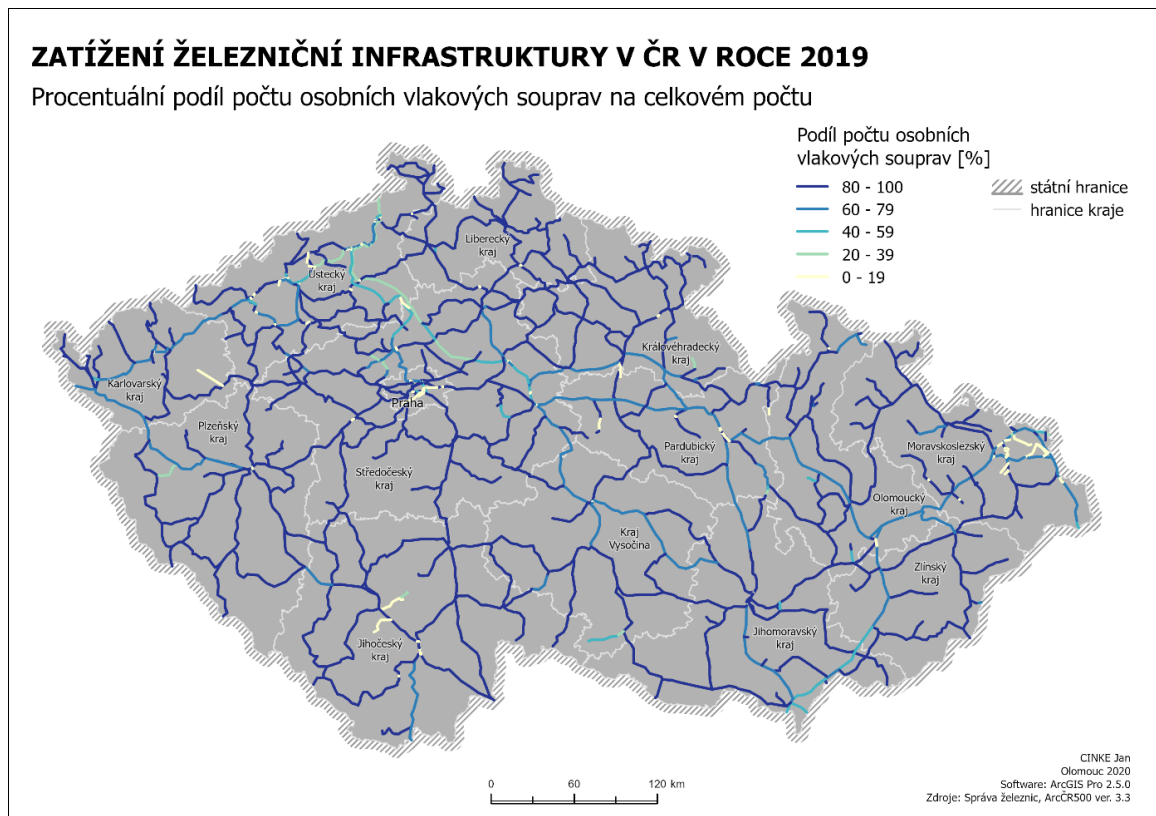
Obr. 5.1 Celkový počet vlakových souprav na úsecích železniční sítě ČR v roce 2019.

Z hlediska hmotnosti (viz obr. 5.2) můžeme vidět, že zmíněné tranzitní koridory ještě více vystupují. Kromě nich jsou vidět další využívané trasy, například úsek Děčín – Nymburk – Kolín – Havlíčkův Brod – Brno.



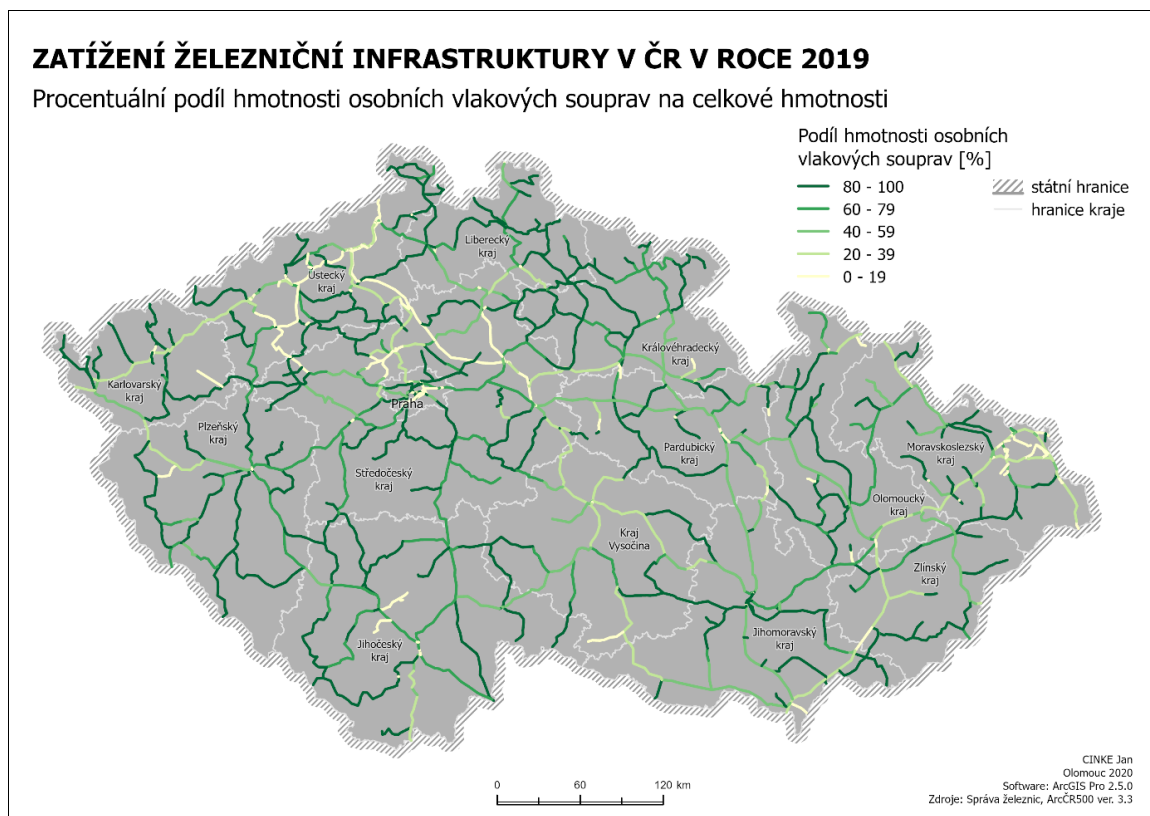
Obr. 5.2 Celková hmotnost vlakových souprav na úsecích železniční sítě ČR v roce 2019.

Dále byly vytvořeny mapové výstupy, kde je zobrazeno procentuální zatížení trati váhou a počtem osobních a nákladních vlakových souprav. Konkrétně šlo o dvě mapy, kde jedna zobrazuje procentuální zatížení osobními vlakovými soupravami na celkové hmotnosti (osobní a nákladní dohromady) a druhá procentuální zatížení počtem osobních vlakových souprav na celkovém počtu. Výsledky jsou rozděleny do 5 kategorií po 20 %. Například pokud je v mapě zobrazena 20% podíl na zatížení hmotností osobními vlakové soupravy, znamená to, že zbytek zatížení do 100 % tvoří vlakové soupravy nákladní. V případě počtu vlakových osobních vlakových souprav (viz obr. 5.3) dominují kategorie 80–100 % a 60–79 %. Jde především o regionální tratě menšího významu. Na hlavních koridorech se v tomto případě přidávají také kategorie 40–59 % a 20–39 % osobních vlakových souprav. To je zapříčiněno větším množstvím nákladních vlakových souprav, vypravovaných na delší vzdálenosti, a tudíž jedoucích po tranzitních koridorech. Poslední kategorie, 0–19 % osobních vlakových souprav, je zastoupena například na krátkých tratích, které obsluhují průmyslové podniky nebo je zde ve velkém prováděna těžba dřeva.



Obr. 5.3 Procentuální podíl počtu osobních vlakových souprav na celkové počtu v ČR v roce 2019.

V případě mapy s procentuální hmotností (viz. obr. 5.4) je vidět rovnoměrnější rozdělení jednotlivých kategorií. To je zapříčiněno tím, že osobních vlakových souprav jezdí násobně více, ale průměrná váha soupravy osobních vlakových souprav nedosahuje váhy souprav nákladních. Osobní vlakové soupravy jsou objednávány na základě poptávky, takže je v regionální dopravě velké množství jedno, nebo dvou vozových souprav, které mají velmi malou hmotnost, ale v celkovém součtu tvoří nezanedbatelnou část z hlediska počtu. Naopak vlakové soupravy nákladní jsou často vypravovány, pokud je to možné, jako dlouhé mnoha vozové soupravy, aby bylo zajištěno co nejmenší blokování infrastruktury a snížení ceny přepravovaného materiálu. Takto zpravidla fungují nákladní vlakové soupravy v mezistátní, státní i regionální dopravě. Největší intenzita nákladní dopravy se soustřeďuje v Moravskoslezském, Ústeckém a Středočeském kraji. V prvních dvou zejména, kvůli blízkosti uhlénné těžby, těžkého průmyslu a blízkosti k hlavním nákladním tranzitním koridorům pro mezistátní přepravu. V případě kraje Středočeského a Prahy zejména proto, že jimi prochází všechny nákladní koridory pro mezinárodní přepravu a jsou zde také umístěny nakládkové kontejnerové terminály. Příkladem může být kontejnerový terminál společnosti Metrans v Pražské Uhříněvsi.

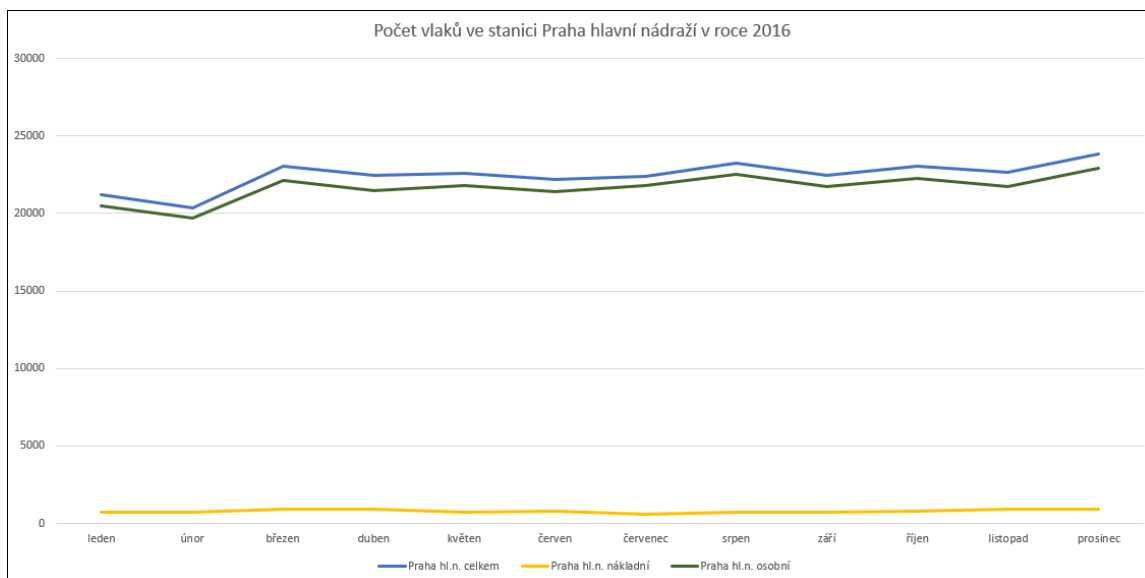


Obr. 5.4 Procentuální podíl hmotnosti osobních vlakových souprav na celkové hmotnosti v ČR v roce 2019.

5.2 Grafy

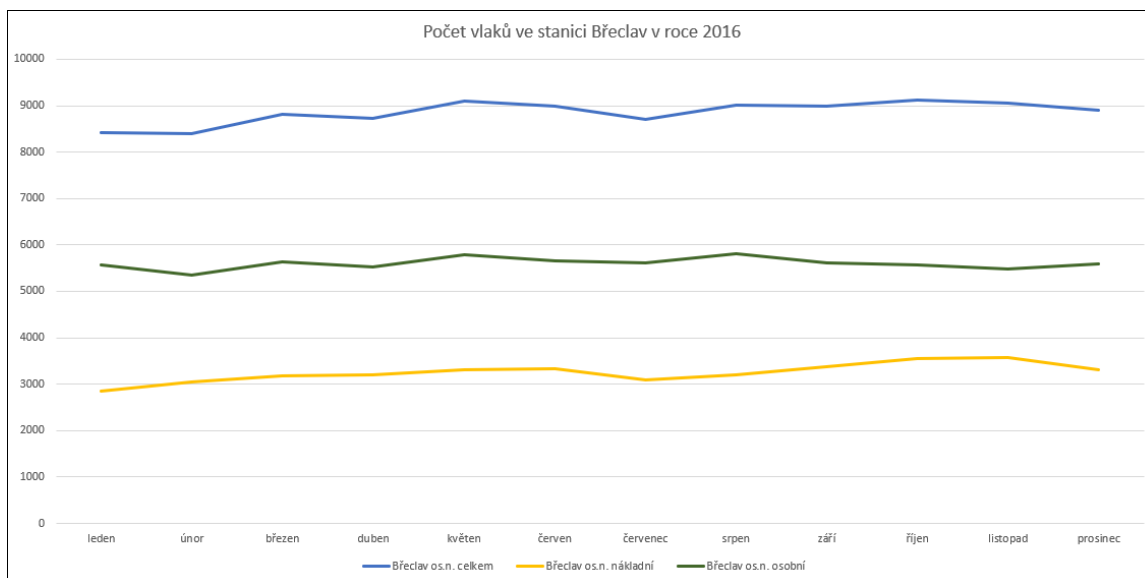
Pro zobrazení variability vstupních dat pomocí grafu měsíčních hodnot za jeden rok byly vybrány data za jednotlivé monitorovací body z roku 2016. Konkrétně byly vybrány stanice Praha hlavní nádraží, Břeclav a Adršpach. Předpoklad při výběru byl, ukázat 3 stanice, které jsou odlišné jak svojí funkcí, tak hodnotami. Stanice Praha hlavní nádraží je typickým představitelem osobního železničního uzlu, mezinárodního významu, ležícího na několika tranzitních železničních koridorech. Břeclav zatupuje železniční uzel ležící v těsné blízkosti hranic s osobní i nákladní dopravou. Posledním případem je Adršpach, který zastupuje regionální stanici s převážně osobním provozem, který je součástí turistické infrastruktury.

Dle předpokladu první ze stanic, Praha hlavní nádraží (viz obr. 5.5), vykazuje primárně provoz osobních vlakových souprav oproti nákladním. Je to dáno tím, že tato stanice je určena primárně pro osobní dopravu, a tudíž je tento jev očekávatelný. Touto stanicí projedou nákladní vlakové soupravy v řádu stovek, až jednotek tisíc souprav za jeden měsíc a jde tedy ve srovnání s osobní dopravou o marginální počet. Tento počet je v rámci roku poměrně neměnný a osciluje od průměrné hodnoty jen nepatrně. Naopak počty osobních vlakových souprav s v rámci roku mění poměrně ve velké míře. Nejdříve je vidět mírný propad v od ledna do února a poté nárůst na poměrně konstantní hladinu od března do listopadu, na závěr můžeme vidět stoupání od listopadu do prosince z důvodu cestování osob na vánoční svátky. Řádově jde o hodnoty mezi 20 000 až 23 000 vlakových souprav za jeden měsíc.



Obr. 5.5 Graf ročního průběhu počtu vlaků ve stanici Praha hl. nádraží v roce 2016.

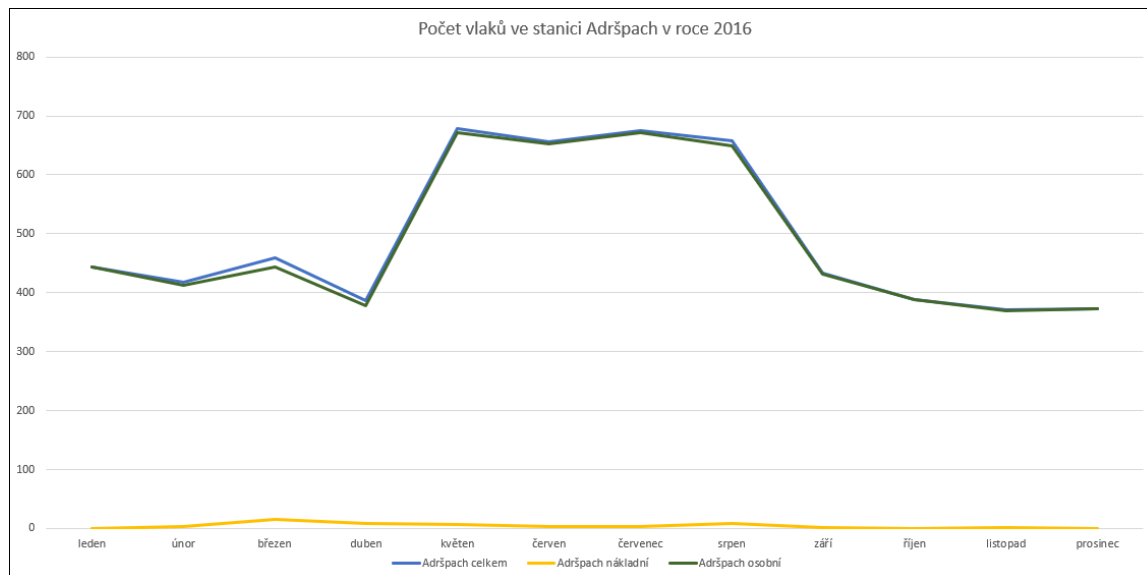
Zcela jiný průběh hodnot můžeme vidět na grafu zobrazujícím hodnoty počtu vlakových souprav ve stanici Břeclav (viz obr. 5.6). Zde je přibližně rovnoměrně rozdělený celkový počet mezi nákladní a osobní dopravu. Průběh počtu osobních vlakových souprav přibližně kopíruje chování osobních vlakových souprav ve stanici Praha hlavní nádraží. Počet osobních vlakových souprav je zde ale diametrálně odlišný, jde přibližně o hodnoty mezi 5 200 až 5 900. Nákladní doprava se, jak je napsáno výše, pohybuje ve zhruba polovičních hodnotách. Konkrétně jde o hodnoty od 2 800 do 3 600. Minimálních hodnot dosahuje nákladní doprava v lednu a poklesy jsou pak také v červenci a prosinci. Mezi těmito hodnotami doprava mírně stoupá.



Obr. 5.6 Graf ročního průběhu počtu vlaků ve stanici Břeclav v roce 2016.

Stanice Adršpach (viz obr. 5.7) má z vybraných stanic asi největší relativní rozptyl. Tento rozptyl je ale pouze v osobní dopravě, ta je zde naprosto převažující nad dopravou nákladní, která se pohybuje v řádech jednotek až desítek za jeden měsíc, některé měsíce zde nákladní doprava vůbec neprobíhá. Osobní doprava se drží od 370 do 680

vlakových souprav za jeden měsíc. Je zde primárně vidět vliv léta a letních prázdnin, kdy jízdni řád obsahuje mnohem více spojů v reakci na množství návštěvníků proudících do Adršpaško-teplických skal. V dubnu začíná doprava stoupat a od května do září se pohybuje v maximálních hodnotách. Následně od srpna do září opět klesá. Zajímavý je také pokles v dubnu před nárůstem letní dopravy.



Obr. 5.7 Graf ročního průběhu počtu vlaků ve stanici Adršpach v roce 2016.

6 ANALÝZA ČASOVÝCH ŘAD

Vzhledem k tomu, že získaná úseková data o využití železniční infrastruktury, byla získána v měsíčních úhrnech za 4 roky, od roku 2016 do roku 2019, bylo vhodné je dále analyzovat z hlediska času. Pro tento úkol je vhodná zejména analýza časových řad, kde bylo možné popsat chování jednotlivých ukazatelů napříč časem, sledovat opakující se vzorce v datech, jejich trend a další prvky. Kvůli velkému množství poskytnutých dat nebylo možné provést analýzu každého monitorovacího bodu nebo úseku na železniční síti v ČR. Konkrétně se jednalo přibližně o 3800 úseků za každý rok. Bylo tedy nutné na základě znalosti dat získané předchozím zkoumáním a vlastní znalostí zmenšit zkoumané množství úseků. Bylo to provedeno vydefinováním význačných kategorií, které mají zajímavý roční průběh, nebo jsou zajímavé z hlediska nějakého fenoménu v železniční přepravě.

6.1 Kategorie

První vybranou kategorií byly tranzitní koridory označené jako kategorie A. Jde o nejvytíženější trasy napříč ČR s regionální, národní i mezinárodní dopravou. Vybrány byly takové úseky, které leží mezi význačnými železničními uzly na již zmíněných tranzitních koridorech. Úseky byly vybrány tak, aby popisovaly chování na všech tranzitních koridorech nebo jejich kombinacích v celé ČR konkrétně šlo o 9 kombinací. Vyjmuty jsou části koridorů nebo jejich kombinace v blízkosti hraničních přechodů, ty budou popsány ve vlastní kategorii.

- Svojkovice z. – Rokycany-Borek z. (3. tranzitní koridor)
- Lukavec z. – Nové Kopisty z. (1. a 4. tranzitní koridor)
- Tábor – Čápův Dvůr z. – Tábor (4. tranzitní koridor)
- Choceň – Brandýs n. O. předj. k. (1. a 3. tranzitní koridor)
- Skalice nad Svitavou – Svitávka z. (1. tranzitní koridor)
- Zábřeh na Moravě – Lupěné z. (3. a 2. tranzitní koridor)
- Hulín – Záhlinice z. (2. tranzitní koridor)
- Jistebník – Studénka (3. a 2. tranzitní koridor)
- Vendryně – Třinec centrum z. (3. tranzitní koridor)

Další a velmi zajímavou kategorií jsou hraniční přechody označené jako kategorie B. Zde je použití úseků velmi vhodné. Protože přes hranice vede vždy v datech pouze jedna linka. Zároveň je úsek vždy vybrán až po státní hranici, takže nejsou uvažovány vlakové soupravy, které končí na území ČR, ale pouze ty vlakové soupravy, které hranice reálně přejíždějí. V této kategorii byly vybrány úseky u všech hraničních přechodů, které leží na tranzitních koridorech.

- Břeclav st. hr – Břeclav os. n.
- Děčín st. hr. – Dolní Žleb
- H. Dvořiště st. hr. – Horní Dvořiště
- Cheb st. hr. – Pomezí náv. Lo, So
- Lanžhot st. hr. – km 9,900 –0,022
- Mosty u Jabl. st. hr. – Mosty u Jabl. zast.
- Petrovice u K. st. hr. – Petrovice u Karviné

Poslední kategorie byla společná pro úseky z hlediska osobní a nákladní dopravy. Vybrány byly takové úseky, ve kterých je zvýšený nákladní provoz nebo zajímavý osobní provoz. Původcem jsou zde především tratě s regionálním osobním provozem a jinak využívané tratě pro nákladní dopravu. Kategorie byla pojmenována jako zajímavé a označena jako kategorie C.

- Mladá Boleslav město – MB hl. n. kol. č. 5c
- Rýmařov – Jamartice z.
- Kouty nad Desnou – Loučná n. Des. -R. nz.
- Adršpach – Teplice n. M. skály z.
- Jemnice – Lhotice u Jemnice z.
- Lednice – Lednice rybníky z.

6.2 Úprava dat

Po výběru kategorií a jednotlivých úseků bylo nutné data dále připravit pro analýzu. Vzhledem k tomu, že lze data rozdělit na osobní a nákladní dopravu, kde některé kategorie přímo s tímto rozdílem pracují, bylo nutné data filtrovat tak, aby bylo možné v analýze tyto rozdíly zobrazit. Pro každou kategorii byly filtrovány jednotlivé vybrané úseky. Data byla vybírána z předpřipravených kontingenčních tabulek za jednotlivé roky v měsíčních agregacích. Při filtraci dat vznikl problém u dvou úseků z kategorie hraniční přechody. U hraničního přechodu H. Dvořiště a Lanžhot došlo v rámci sledovaného období ke změně druhého z dvojce monitorovacích bodů, které definují tyto úseky. Muselo tedy dojít ke sloučení hodnot z nového a starého označení tohoto úseku a v konfliktní měsíc musely být sečteny hodnoty z obou těchto úseků. Vznikly tabulky s vybranými úseky za jednotlivé kategorie rozdělené na osobní a nákladní dopravu.

Dále bylo nutné, už v rámci kroků nutných v analýze časových řad, zamezit nesouměřitelnosti dat, vzniklé rozdílnou délkou měsíce. Existuje více metod, které ve své publikaci popisuje například Hančlová (2003), které lze pro vyřešení tohoto problému použít. Vzhledem k tomu že chceme porovnávat měsíční časovou řadu ze

čtyř let, je vhodné, aby vyrovnaní zohlednilo souměřitelnost jednotlivých měsíců, nejen v rámci jednoho roku, ale i v rámci celého časového období. Vznikla, tak souměřitelná časová řada, kde bylo možné porovnat stejné časové období mezi roky 2016 a 2019 včetně. Na základě požadavků definovaných výše bylo vybráno vyrovnaní na ekvivalentní délku měsíce pro všechny 4 roky a to 30 dní. Přepočítání proběhlo na základě následujícího vzorce.

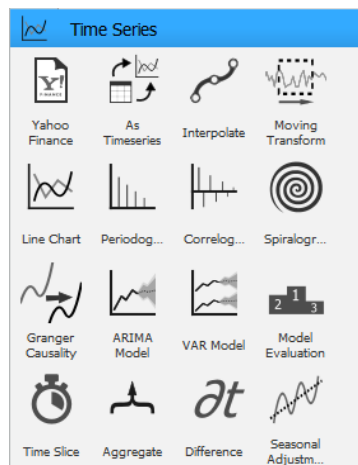
$$y_t^{(\text{očistěná})} = y_t \times \frac{p}{p_t}$$

Kde:

- $y_t^{(\text{očistěná})}$ = očistěná hodnota od vlivu rozdílné délky měsíce,
- y_t = původní hodnota zatížená rozdílnou délkou měsíce,
- p = fixní délka měsíce 30 dní,
- p_t = počet dní v měsíci t .

6.3 Dekompozice časových řad

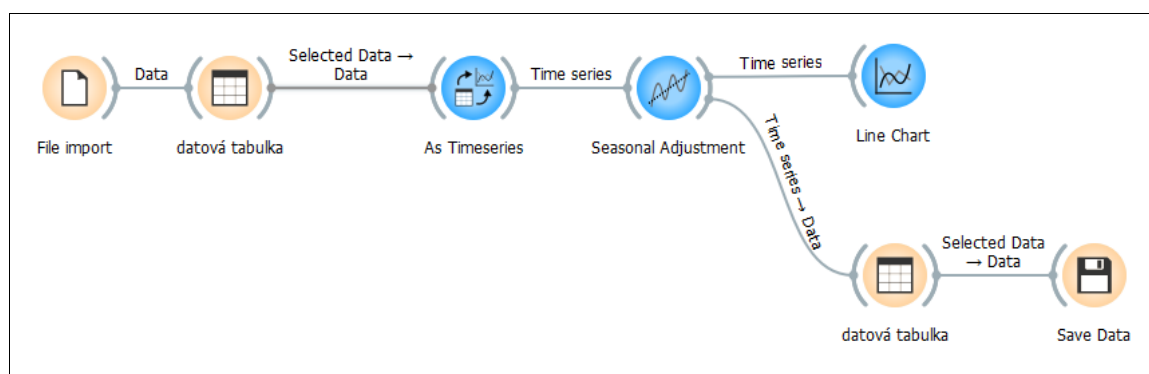
Pro dekompozici časových řad byl použit software Orange. Konkrétně jde o doplněk *Time Series*, který je dostupný jako rozšíření. Instalace probíhá přes volbu *Options*, na horní liště programu a poté volbu *Add-ons*. Poté už byl jen vybrán doplněk *Orange3 Timeseries*. Po instalaci se v programu zobrazila, kromě již defaultně nastavených karet, nová karta *Time series* (obr. 6.1). Na této kartě jsou všechny dostupné nástroje pro zpracování časových řad, které jdou kombinovat s ostatními nástroji z jiných karet.



Obr. 6.1 Karta Timeseries v programu Orange.

Pro dekompozici byl vybudován v programu Orange model (viz obr. 6.2) pomocí již zmíněných nástrojů z karty Timeseries. Nejdříve byl zařazen nástroj pro nahrání dat, data byla nahrána ve formátu CSV a ve struktuře, 48 řádků s informacemi pro každý měsíc ze 4 let, pro které byla data dostupná a sloupce pro jednotlivé vybrané úseky. Poté pro kontrolu dat v tabulce byla použita funkce *tabulka*, která zobrazuje data. Dále byl vybrán, už z karty *Timeseries*, nástroj *As Timeseries*, který znovu interpretuje jakoukoli datovou tabulku jako časovou řadu, takže ji lze použít se zbytkem nástrojů

v doplňku *Timeseries*. Tento nástroj má pouze jeden parametr, který definuje posloupnost časové řady. Buďto jde vybrat příslušný atribut nebo zaškrtnou, že pořadí záznamů v tabulce odpovídá pořadí posloupnosti časové řady. Dalším nástrojem zvoleným v tomto modelu je nástroj *Seasonal Adjustment*. Tento nástroj rozloží původní časovou řadu, která musí být výstupem z nástroje *As Timeseries*, na jednotlivé složky, jimiž jsou sezónní složka, trend, reziduální složka a sezónně upravená časová řada. V tomto nástroji mohou být vybrány dva parametry a časové řady, které chceme rozložit. Prvním parametrem je délka sezónního cyklu, v tomto případě jde o jeden rok, tedy 12 měsíců. Poté může být nastaveno, zda chceme provést rozklad aditivní nebo multiplikativní. Lze pouze o to, zda se výsledné složky sčítají, v případě aditivního, nebo násobí, v případě multiplikativního. V této práci byl zvolen aditivní rozklad. Pro zobrazení rozkladu časové řady byl použit nástroj line chart, který je jednoduchou možností zobrazit všechny složky pohromadě v jednom grafu. Vzhledem k malým možnostem programu Orange, zejména z hlediska práce s grafy byl ještě do modelu přidán nástroj pro export výsledných dat tak, aby bylo možné data dále zpracovávat a sjednotit jejich výslednou podobu.



Obr. 6.2 Model pro rozklad časových řad v programu Orange.

6.4 Hodnocení

V této kapitole budou porovnány, na základě výstupů vzniklý rozkladem časových řad, rozdíly v rámci jednotlivých kategorií a druhů dopravy. Konkrétně půjde například o porovnání hraničních přechodů z hlediska nákladní dopravy. Pro porovnání průběhu hodnot v rámci celého časového období, budou použity grafy s vynesnými hodnotami počtu osobní a nákladní dopravy vyrovnanými na ekvivalentní délku měsíce 30 dní, aby nebyly počty zkreslené. Pro vývoj hodnot bez vlivu reziduálních výkyvů bude použit graf trendů pro jednotlivé úseky a typy dopravy. Analýza časové složky poskytnutých dat si nedává za cíl primárně popisovat rozdíly mezi osobní a nákladní dopravou. Tato problematika je popsána v následující kapitole (kap. 7), kde se na jejím základě provádí kategorizace.

Vzhledem k nečitelnosti grafů použitých v textu, budou grafy v dostatečné velikosti umístěny v příloze (příloha 1-3) a v kapitolách budou pouze grafy v menší podobě pro ilustraci průběhu časové řady, trendu a sezónní složky. V kapitole také nebudou umístěny všechny dostupné grafy, ale pouze vybrané z hlediska subjektivní důležitosti v daném tématu. Ostatní, jak je popsáno výše, budou dostupné v příloze. Vzhledem k rozsahu příloh bylo rozhodnuto použít jako tištěné přílohy pouze grafy pro tranzitní koridory, jakožto reprezentanty struktury příloh k interpretacím časových řad. Grafy pro zbylé kategorie budou umístěny v elektronických přílohách. Číslování příloh bude

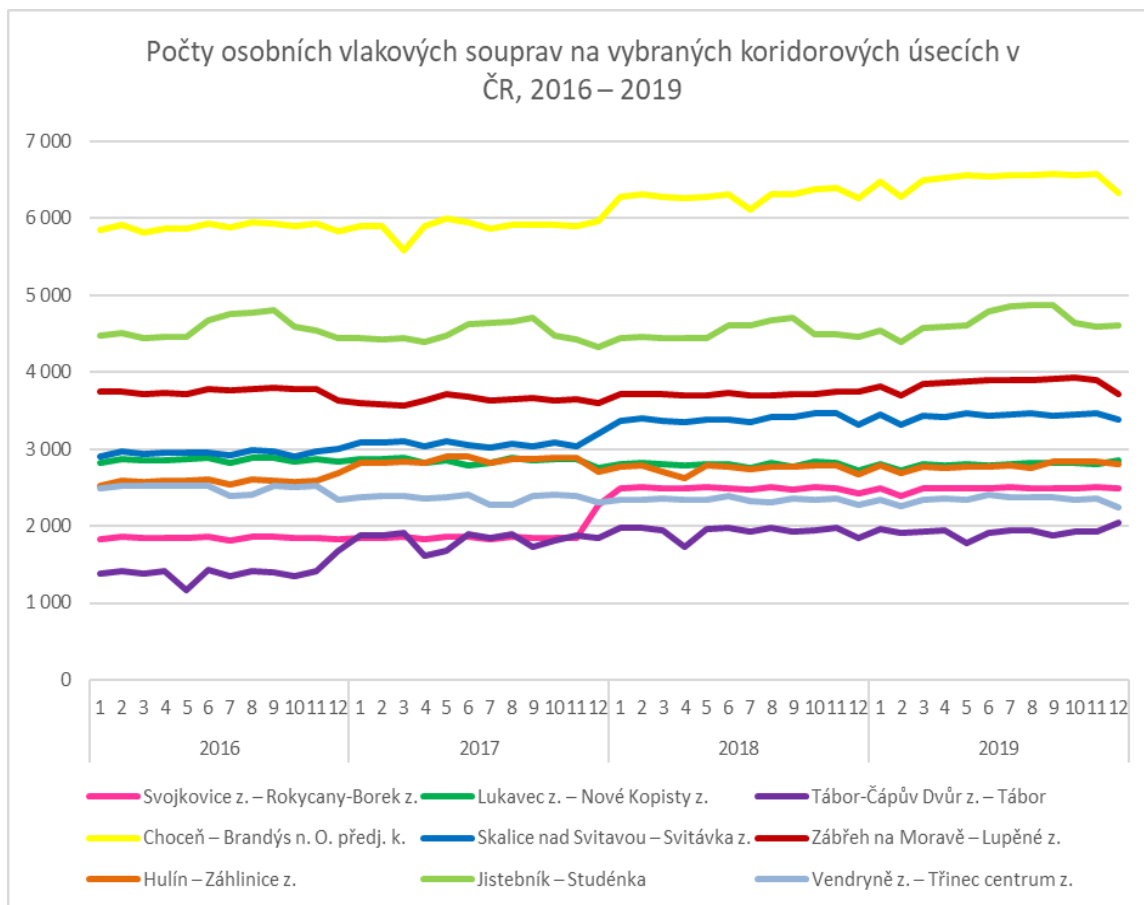
hierarchicky zachováno bez ohledu na to, zda je příloha tištěná nebo elektronická, pro lepší orientaci. V rámci popisu jednotlivých úseků je používáno pojmenování po jedné ze stanic tvořících daný úsek.

6.4.1 A. Tranzitní koridory

Většina úseků z této kategorie se z hlediska počtu osobních vlakových souprav pohybuje, s přibližně stejnými rozestupy, v hladině 1 400 až 3 900 vlakových souprav za měsíc (viz obr. 6.3). Změna nastává u úseků Choceň, Zábřeh na Moravě a Jistebník. Všechny tyto úseky leží na 3. tranzitním koridoru. Jistebník a Zábřeh na Moravě dosahují o něco vyšších hodnot než většina úseků. Choceň má z vybraných úseků absolutně nejvyšší hladinu počtu vlakových souprav za měsíc. Největší volatilitu, v měřítku měsíčních dat, vykazují úseky Choceň a Jistebník. Skokový nárůst hladiny počtu osobních vlakových souprav je zřetelný v prosinci 2016 u úseku Tábor – Čápův Dvůr a v prosinci 2017 u úseků Svojkovice a Skalice nad Svitavou. Měsíční propady jsou převážně způsobeny plánovanou výlukovou činností a nepravidelnými mimořádnostmi. Přibližně stacionární trend vykazují úseky Jistebník, Lukavec a Vendryně. Částečně nebo úplně stoupající trend vykazuje zbytek stanic. Zde je individuální, zda je růst po celou dobu nebo jen v části sledovaného období. V některých případech je také vidět nepatrný pokles. Pouze klesající trend nevykazuje žádný z vybraných úseků, to znamená, že osobní doprava na hlavních koridorech spíše stoupá po celé sledované období. Z dat jasně vystupuje zatíženost 3. tranzitního koridoru mezi Prahou a Ostravou. Nejmenší zatížení naopak vykazuje 4. tranzitní koridor v úseku z Prahy na České Budějovice.

Z hlediska popisu sezónní složky jsou popisovány úseky se stacionárním vývojem trendu. V tomto případě tedy Jistebník, Lukavec a Vendryně. Úsek Jistebník vykazuje dvě úrovně chování během periody sezónní složky, nižší hladinu a vyšší hladinu. Nižší hladina trvá od listopadu do května a vyšší od června do října. Maximální hodnoty vykazuje v září. Úsek Lukavec má oproti předchozímu mnohem menší výkyvy od střední hodnoty sezónní složky. V podstatě celá perioda je vyplněna oscilací kolem střední hodnoty s minimem v prosinci a maximem v srpnu. Úsek Vendryně je nejdříve od ledna do května poměrně blízký svými hodnotami střední hodnotě sezónní složky, poté narůstá v červnu a pak klesá na červenec a srpen pod střední hodnotu. Poté stoupá nad střední hodnotu od září do listopadu a pak opět klesá. Z hlediska prázdninových měsíců vykazuje úsek Vendryně opačné chování než Jistebník a je zajímavé, že vliv prázdninových měsíců není vždy totožný.

Při celkovém pohledu je zřetelný zejména nárůst hladiny počtu osobních vlakových souprav v úsecích Svojkovice na 3. tranzitním koridoru Praha – Plzeň, Skalice nad Svitavou na 1. tranzitní koridoru Česká Třebová – Brno a Tábor na 4. Tranzitním koridoru Praha – České Budějovice. Dále pak každoroční nárůst dopravy v úseku Jistebník na 2. a 3. tranzitním koridoru v letních měsících, kdy v ostatních úsecích nebyla reakce na toto období zaznamenána, nebo byl zaznamenán mírný pokles. Obecně je u osobní dopravy až na výjimky typická malá volatilita v rámci měsíců, daná platným jízdním řádem.



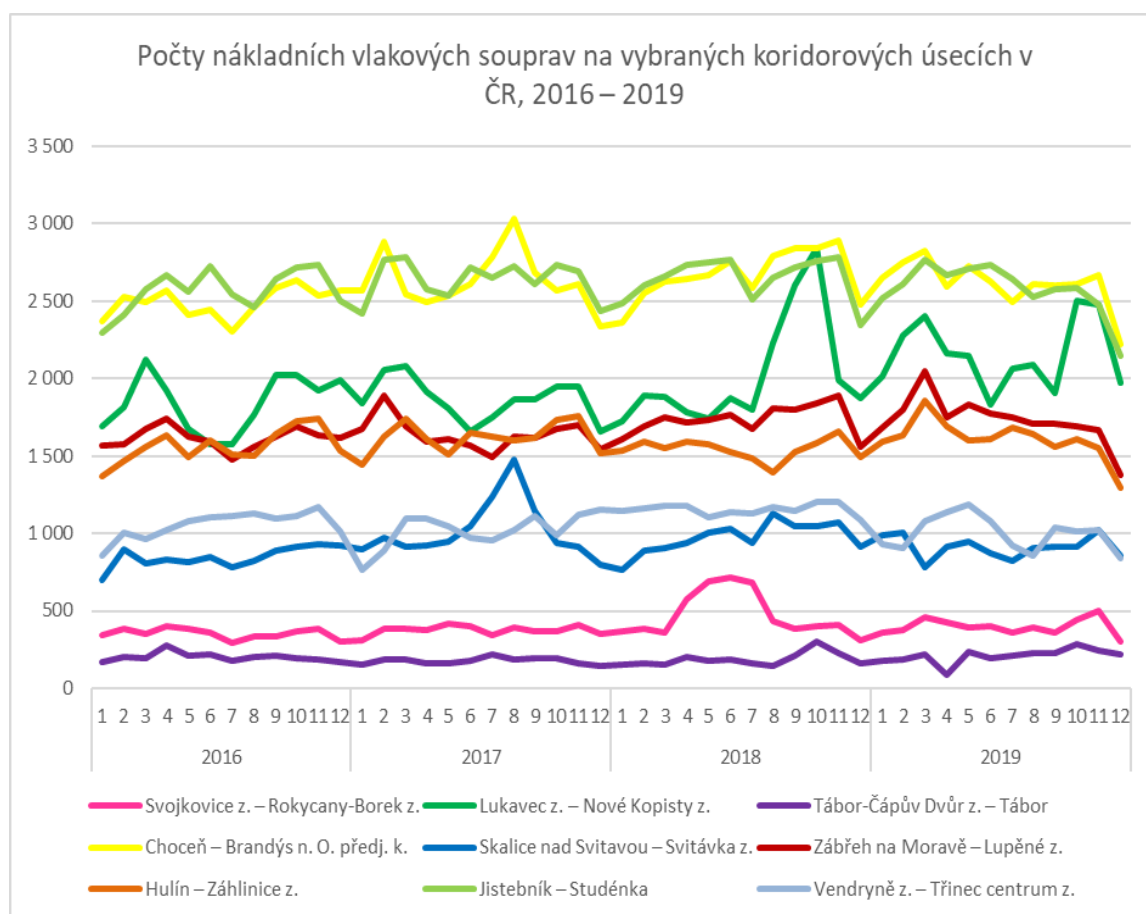
Obr. 6.3 Počet osobních vlakových souprav na vybraných koridorových úsecích v ČR, 2016–2019.

Při prvním pohledu na průběhy hodnot časových řad vybraných koridorových úseků pro počet nákladních vlakových souprav (viz obr. 6.4) je zřetelné, že mnohem více kolísají než časové řady pro počet osobních vlakových souprav. Tato skutečnost je dána především částečnou nepravidelností nákladní vlakové přepravy, jejíž přepravní výkon je dán poptávkou, nikoli pravidelným jízdním řádem jako u dopravy osobní. Dle průměrné hladiny počtu vlakových souprav za měsíc pro jednotlivé úseky můžeme křivky rozdělit do 4 kategorií. Nejmenších počtů dosahují úseky Tábor a Svojkovice. Vyšších počtů poté úseky Skalice nad Svitavou a Vendryně. Ještě vyšších počtů dosahují úseky Lukavec, Hulín a Zábřeh na Moravě. Nejvyšších hodnot poté dosahují úseky Choceň a Studénka. Za celé sledované období čtyř let vykazuje většina úseků poměrně nestabilní trend. Stacionární je pouze chování trendu v úseku Tábor. Dále, pokud by mělo dojít k zobecnění, tak trend s počátečním nárůstem a poté poklesem na hodnotu vyšší, než byla počáteční, vykazují úseky Choceň, Jistebník, Svojkovice a Skalice nad Svitavou. Úseky Lukavec, Hulín a Zábřeh na Moravě vykazují proměnlivý trend s koncovou hodnotou vyšší, než je ta počáteční. Celkový pokles zaznamenal pouze úsek Vendryně, který nejdříve v první části do května 2017 klesal, poté stoupal do května 2018 a následně do konce sledovaného období klesal pod počáteční úroveň.

Z hlediska sezónní složky byl vybrán pouze úsek Tábor, protože jako jediný vykazoval stacionární trend, a tudíž byl vhodný pro interpretaci opakujícího se sezónního chování. Od počátku stoupá křivka lehce nad střední hodnotu v březnu. Následně klesá do minima celé sezónní složky v dubnu. Poté stoupá ke střední hodnotě a od května do srpna kolem ní osciluje. Následně stoupá ke svému maximu v říjnu a odtud strmě klesá

k prosinci. Lze tedy říci, že od února do března a poté od května do srpna vykazuje sezónní složka poměrně neutrální hodnoty. Výchylinky přicházejí od listopadu do února a od března do května směrem k nižším hodnotám. Výchylnka směrem k vyšším hodnotám přichází od srpna do listopadu.

Z celkového pohledu nelze jednoznačně říci, jestli se v rámci roku vyvíjí počty nákladních vlakových souprav stejným způsobem. Podobnosti jsou zřetelné při pohledu na úseky, které leží alespoň z části na stejných tranzitních koridorech. Opět je zde vidět vedoucí pozice hodnot na koridoru Praha – Ostrava. Oproti hodnotám osobních souprav však vystupuje koridor Praha – Děčín, který se v některých měsících přibližuje úhrnu počtu nákladních souprav k úsekům Choceň a Jistebník. Úsek Lukavec, také obsahuje poměrně velkou výchylnku, zejména v letech 2018 a 2019 od srpna do listopadu, kdy se nákladní doprava zvýší o přibližně 1 000 až 1 500 vlakových souprav.



Obr. 6.4 Počet nákladních vlakových souprav na vybraných koridorových úsecích v ČR, 2016–2019.

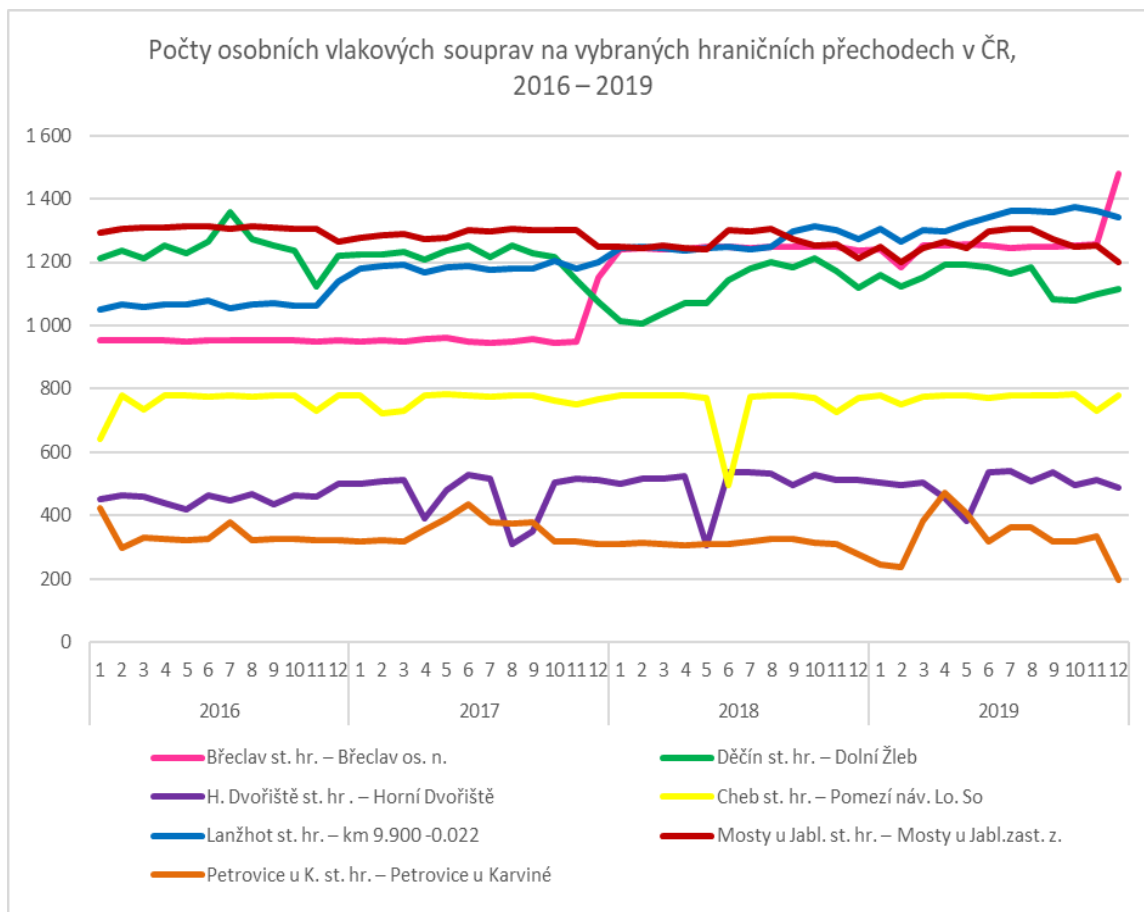
6.4.2 B. Hraniční přechody

V rámci osobní dopravy je už na první pohled zřejmé, že se úseky pohybují v rozdílných absolutních hodnotách počtu osobních vlakových souprav. Vizualně lze tyto hodnoty rozdělit na tři kategorie (obr. 6.5). Nejnížší kategorie, ve které se nachází Horní Dvořiště a Petrovice u Karviné. Prostřední kategorie, ve které se nachází Cheb a nejvyšší kategorie, kde se nachází zbytek úseků z kategorie B. Toto rozdělení je ještě více zřetelné na zobrazení trendu počtu osobní dopravy. Všechny velké výkyvy v rámci počtu osobní dopravy jsou především tvořeny plánovanými výlukami a neplánovanými

mimořádnostmi. Hraniční přechody Horní Dvořiště, Petrovice u Karviné, Cheb a Mosty u Jablunkova vykazují přibližně stacionární trend vývoje počtu osobní dopravy. Další hraniční přechody už vykazovaly rozdílné chování trendu. Jinak stacionární hraniční přechod Břeclav zaznamenal skokový nárůst počtu vlakových souprav mezi listopadem 2017 a lednem 2018, což je zřetelné i na grafu trendu. Hraniční přechod Děčín jako jediný hraniční přechod vykazoval klesající trend počtu osobních vlakových souprav, navíc ještě s mimořádným výkyvem směrem k nižšímu počtu v dubnu 2018. Hraniční přechod Lanžhot jako jediný zaznamenal zřetelný nárůst za celé sledované období, který je po celou dobu kontinuální.

Pro popis sezónní složky byly vyřazeny ty časové řady, jejichž trend nebyl stacionární a ty řady, které obsahovaly velké výkyvy z důvodu mimořádností, protože příliš ovlivňovali výslednou křivku. Pro popis opakujícího se chování s periodou jeden rok zbyly hraniční přechody Mosty u Jablunkova a Petrovice u Karviné. Sezónní složka Mostů u Jablunkova se dá rozdělit do dvou částí, první je období od prosince do května, kdy je křivka v nižších hodnotách a poté od června do listopadu, kdy je křivka ve vyšších hodnotách. Petrovice u Karviné jsou v nízkých hodnotách od prosince do února. Od února do dubna je v grafu vidět velký skok a poté pozvolný pokles směrem k prosinci. V obou případech je možné říci, že minima v chování sezónní složky dosahuje v zimních měsících a maximálních hodnot v měsících letních.

Z hlediska průběhu křivek za 4 roky je zajímavé především chování hraničního přechodu Břeclav, který na konci roku 2017 prodělal poměrně velký nárůst, a na konci sledovaného období ke konci roku 2019 začala křivka opět růst ze stacionárního stavu. Tato skutečnost se jeví jako pravidelné zvyšování hladiny dopravy s periodou 2 roky.



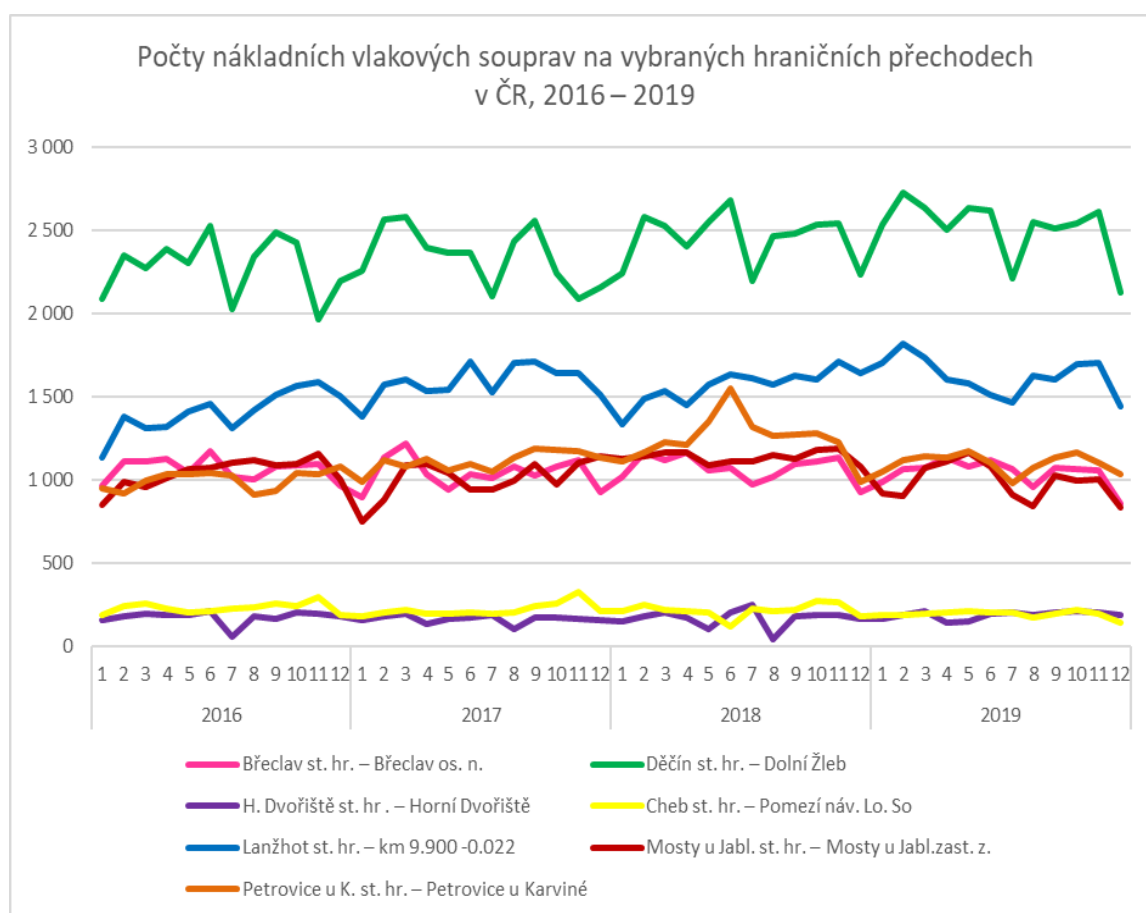
Obr. 6.5 Počet osobních vlakových souprav na vybraných hraničních přechodech v ČR, 2016-2019.

Nákladní doprava vykazuje podobné chování z hlediska rozdílu absolutních hodnot měsíčních úhrnů počtu vlakových souprav (viz obr. 6.6). Opět lze na jejich základě provést rozdělení. Do skupiny s nejnižšími hodnotami spadají hraniční přechody Cheb a Horní Dvořiště. Vyšší kategorie je zastoupena úseky Břeclav, Petrovice u Karviné a Mosty u Jablunkova. Dále už zbývají osamocené hraniční přechody z hlediska počtu vlakových souprav. Tou s nižšími hodnotami je Lanžhot a s vyššími hodnotami Děčín. Obecně vykazuje nákladní doprava více proměnlivé chování v rámci jednotlivých měsíců než doprava osobní. Rekordní je v tom případě hraniční přechod Děčín, který vykazuje volatilitu více než 500 vlakových souprav za měsíc. Zároveň jde o nejvyšší nákladní tranzit do a ze sousedních zemí ze všech sledovaných hraničních přechodů. Z hlediska interpretace grafu trendu lze považovat za stacionární hraniční přechody Břeclav, Horní Dvořiště a Cheb. Rostoucí trend vykazují hraniční přechody Lanžhot a Děčín. Proměnlivý trend mají hraniční přechody Mosty u Jablunkova a Petrovice u Karviné. Mosty u Jablunkova mají nejdříve mírně klesající trend od začátku sledovaného období po květen 2017, poté rostoucí do června 2018 a poté opět klesající. Petrovice u Karviné mají stoupající trend od začátku sledovaného období do května 2018 a poté trend klesající do konce sledovaného období.

Pro interpretaci sezónní složky byly vybrány úseky Horní Dvořiště, Cheb a Břeclav, protože vykazují stacionární trend za celé sledované období a až na výjimky nejsou v jejich průběhu žádné významné výkyvy. Sezónní složka hraničního přechodu Horní Dvořiště v rámci roku osciluje kolem střední hodnoty sezónní složky s vyššími hodnotami v březnu, červnu a říjnu a s nižšími hodnotami v květnu, srpnu a lednu.

Hraniční přechod Cheb se od prosince do srpna pohybuje pod střední hodnotou sezónní složky s minimem v červnu. Od září stoupá nad střední hodnotu, až do listopadu odkud už zase zpět klesá pod střední hodnotu. Sezónní složka hraničního přechodu Břeclav má ze sledovaných největší volatilitu. Lokálních maximálních hodnot dosahuje v březnu, červnu a listopadu s nejvyšší hodnotou v březnu. Lokální minima dosahuje v květnu, červenci a prosinci s minimem v prosinci. Právě hraniční přechod Břeclav je z hlediska průběhu sezónní složky velmi zajímavý. Doprava nabývá silných měsíců na jaře a na podzim, letní měsíce a přelom roku jsou slabší.

Z celkového hlediska je velmi zajímavé chování hraničního přechodu Děčín, který zcela dominuje z hlediska absolutních počtů a také co se týče volatility v rámci jednotlivých měsíců. Dále pak pokles trendu tranzitu v Moravskoslezském kraji od května 2018, kdy pokles zaznamenali Mosty u Jablunkova i Petrovice u Karviné. V rámci trendu je také pozoruhodný téměř neměnný průběh trendu na hraničních přechodech Horní Dvořiště a Cheb.



Obr. 6.6 Počet nákladních vlakových souprav na vybraných hraničních přechodech v ČR, 2016-2019.

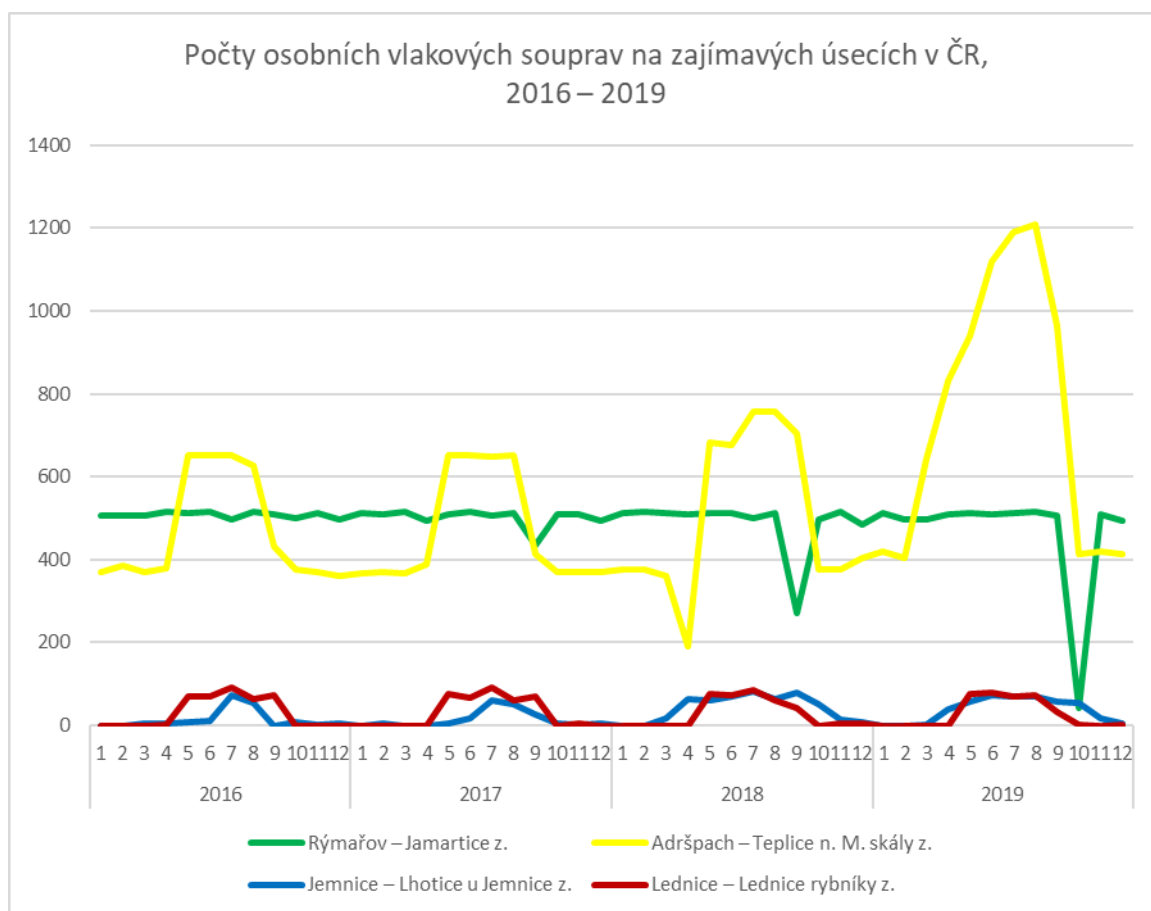
6.4.3 C. Zajímavé

Oproti předchozím dvěma kategoriím bylo při výběru úseků počítáno s jistým předpokladem zajímavého průběhu v rámci časové řady. Výsledné časové řady jsou tedy co se týče chování v rámci jednoho roku, ale i po sobě jdoucích rocích, poměrně dobře interpretovatelné. Jednotlivé úseky byly popisovány na základě stejných ukazatelů jako předchozí dvě kategorie s rozdílem, že bude osobní a nákladní doprava popisována

dohromady. Zobrazené grafy byly rozděleny na úsek Mladá Boleslav (viz obr. 6.9) město a ostatní (viz obr. 6.7 a obr. 6.8), protože úsek Mladá Boleslav nabývá příliš odlišných hodnot pro zobrazení do jednoho grafu.

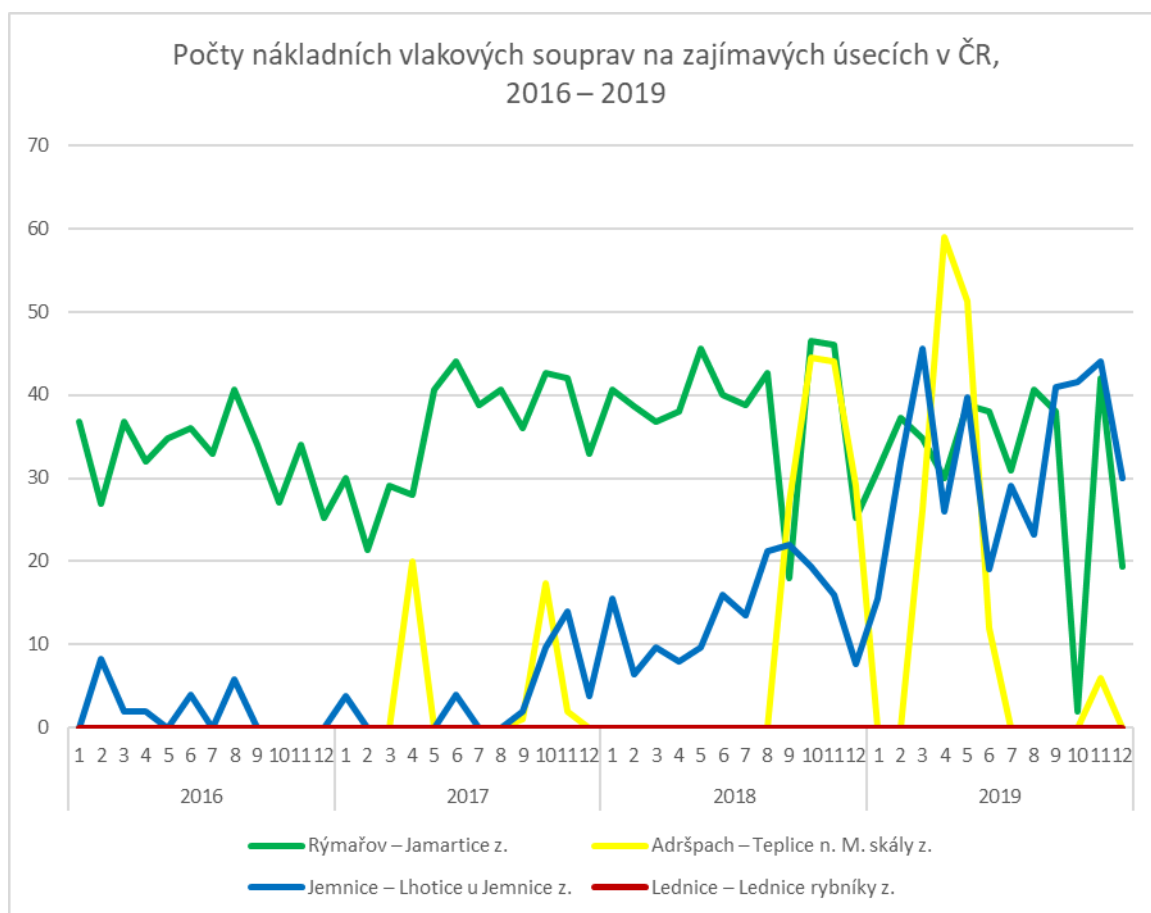
První vybraným úsekem je Rýmařov. Zde už na první pohled zaujme stále rostoucí propad v září v letech 2017 a 2018 a v listopadu v roce 2019 u osobní dopravy a poslední dva jmenované propady i u dopravy nákladní. Tyto propady nejsou zapříčiněny výlukovou činností ani mimořádnostmi na části trati 310 Valšov – Rýmařov. Tento jev je v rámci chování osobní dopravy tím nejzajímavějším. Nákladní doprava vykazuje více proměnné chování. Zde je velmi individuální každý měsíc a nelze přesně odhadnout nějaké přesné chování v rámci roku. S jistotou lze říci, že hodnoty nákladní dopravy od srpna 2018 jsou mnohem více proměnné než před tímto měsícem. Z hlediska trendu vykazuje osobní doprava do března 2019 stacionární chování a po něm pozvolný pokles. Nákladní doprava klesá do října 2016, poté stoupá do listopadu 2017 a poté už klesá do konce sledovaného období. Z hlediska nestacionárního trendu nelze interpretovat graf sezónní složky, ale z původního grafu průběhu počtu osobních vlakových souprav lze vidět jednoznačný propad v září, respektive říjnu. Tento jev je zřetelný u nákladní, ale i osobní dopravy.

Jedním s dalších úseků je Adršpach. Zde je na první pohled zřetelná pravidelnost sezónní osobní dopravy. Jde přibližně o období od dubna do října a každý rok je hladina této letní dopravy vyšší. V roce 2019 je pak letní doprava téměř dvakrát vyšší než v roce předchozím a lze jí přisuzovat velkému zájmu o tuto turistickou lokalitu. Jediný zřetelný propad v ustáleném chování je v dubnu 2018, kdy byla na trati 047 Trutnov – Teplice nad Metují plánovaná výluka v různých úsecích tak, že ovlivňovala počet vlakových souprav projíždějících sledovaným úsekem. Nákladní doprava je v tomto úseku spíše sporadická, ale přesto relativně početná na tuto lokalitu. V jejím výskytu nelze vypořádat nějaké vzory v chování v rámci roku, lze spíše o nárazový nárůst v rámci třech až pěti měsíců, kdy počet stoupne a zase klesne. Největší takovýto skok byl, zaznamenaný mezi únorem a červencem 2019 a dosahoval v dubnovém maximu téměř 60 vlaků za měsíc. Trend osobní dopravy je od začátku sledovaného období lineární až do dubna 2018, kde přechází v rostoucí. Trend nákladní dopravy je díky častým nulovým hodnotám těžko interpretovatelný. Sezónní složka je však zřejmá už ze základního průběhu časové řady a je to již zmiňovaný prázdninový nárůst osobní dopravy od dubna do října.



Obr. 6.7 Počet osobních vlakových souprav na zajímavých úsecích v ČR, 2016-2019.

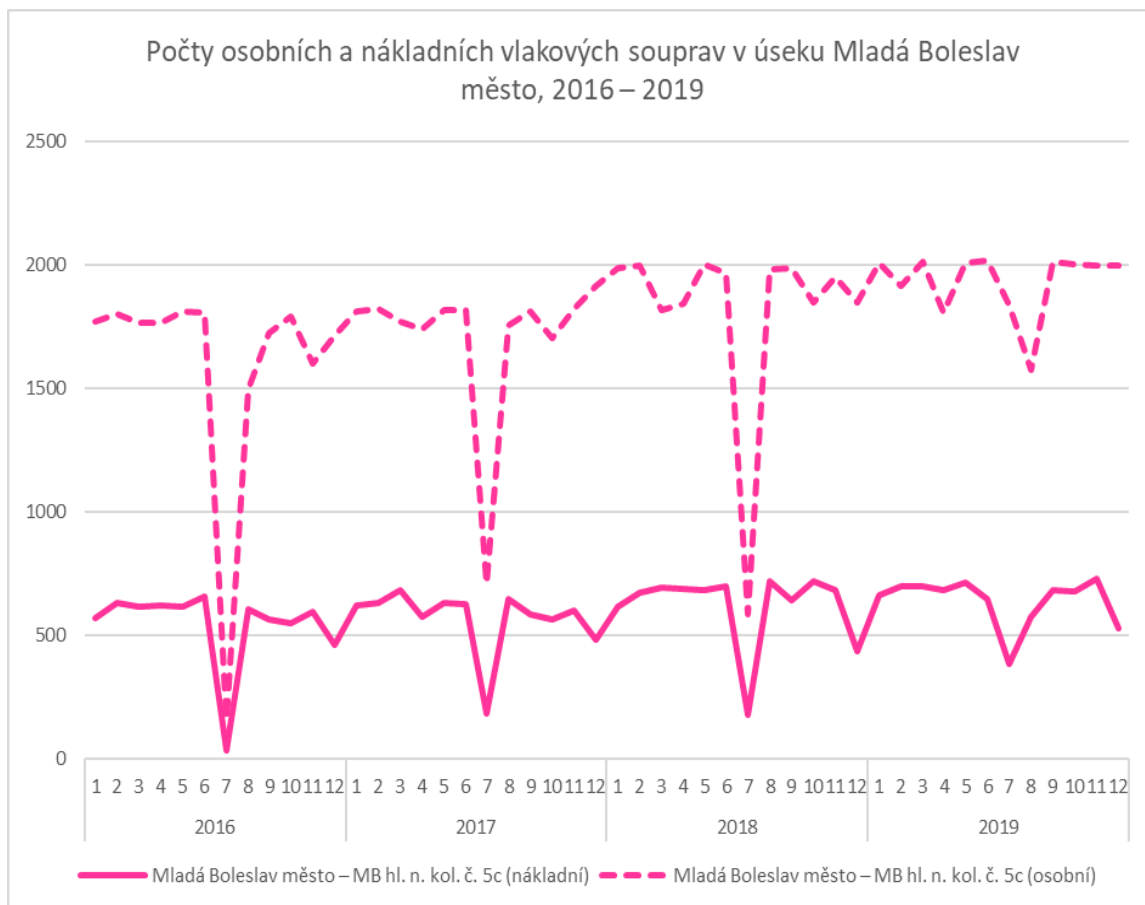
Další zajímavým úsekem z hlediska počtu vlakových souprav je Jemnice. Zde je velmi zajímavý provoz osobní, ale i nákladní doprava. Osobní přeprava vykazuje každoroční sezónní provoz od května do září v roce 2016 až po rozsah od března do listopadu v roce 2019. Měsíční úhrny rostou velmi málo, ale rozsah sezónního provozu se zvětšuje. Jde především u turistický víkendový provoz. V rámci nákladní dopravy je situace odlišná. Měsíční počty kolísají do září 2017 mezi 0 až 8 vlaků za měsíc. Od října 2017 do konce roku se doprava zvyšuje až na 20 vlaků za měsíc. Na přelomu roku 2018 a 2019 se doprava výrazně zvyšuje na hodnoty až 45 vlaků za měsíc. To je zapříčiněno především svozem kůrovcového dřeva z oblasti. Trend osobní dopravy je stacionární, s malým zvýšením hladiny na přelomu roku 2017 a 2018. Trend nákladní dopravy je nejdříve klesající do března 2017 a poté strmě rostoucí do konce sledovaného období. Na základě výše zmíněných průběhů trendů osobní a nákladní dopravy lze pracovat z hlediska sezónní složky pouze s grafem sezónní složky osobní dopravy. Tento graf už pouze doplňuje graf průběhu samotné časové řady, kdy je zřetelně vidět sezónní nárůst přibližně mezi březnem a listopadem. V jiných měsících na trati osobní doprava neprobíhá.



Obr. 6.8 Počet nákladních vlakových souprav na zajímavých úsecích v ČR, 2016-2019.

Další zajímavým úsekem je Lednice. Jde o typickou trať se sezónní osobní dopravou. Nákladní doprava zde neprobíhá. Každý rok je doprava velmi pravidelná od dubna do října s velmi podobným průběhem v rámci měsíců. Ostatní měsíce zde osobní doprava neprobíhá. Pouze v prvních dvou letech v aktivních měsících hodnoty více kolísají než v následujících dvou. Tato trať není typickým představitelem relativně vysokých přepravních výkonů díky letní turistické sezóně, zde je železnice přímo zahrnuta jako prvek turistického ruchu dotvářející Lednicko-Valtický areál a je provozována historickými vlakovými soupravami v poměrně malých počtech. Trend je zde stacionární a sezónní složka kopíruje každoroční průběh časové řady, protože se zde nevyskytují žádné další výkyvy.

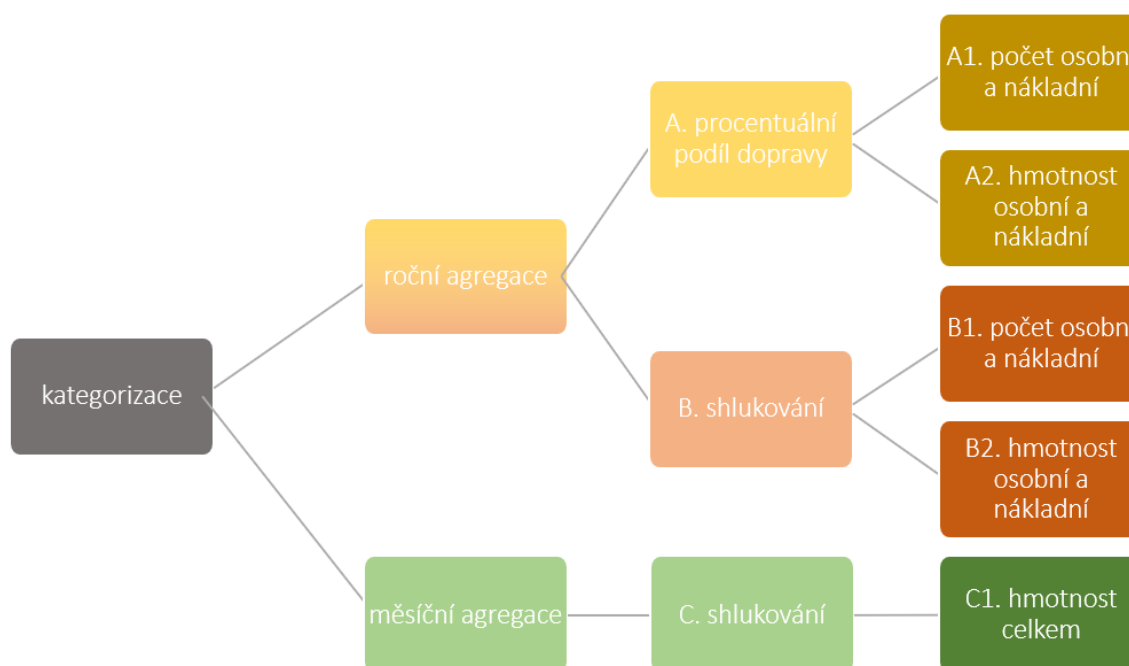
Jako poslední zajímavý úsek byla vybrána Mladá Boleslav město. Jde o úsek, kterým musejí projet všechny nákladní vlaky obsluhující mladoboleslavskou továrnu Škoda auto a zároveň osobní vlaky při vlakové dojíždce za prací do této továrny. Chování časové řady počtů osobních vlakových souprav vykazuje jednu hlavní charakteristickou vlastnost, a to razantní pokles v době celozávodní dovolené v továrně Škoda auto, které je vždy přibližně plánována na červenec a srpen v době trvání 2–3 týdnů. Počty nákladních vlakových souprav zaznamenávají stejný propad, a ještě navíc každoroční propad v prosinci, který však už není natolik velký jako ten způsobený celozávodní dovolenou. Trend osobní i nákladní dopravy v tomto úseku je rostoucí, tudíž nelze interpretovat graf sezónní složky. Sezónní chování je však zřetelné už při pohledu na původní časovou řadu.



Obr. 6.9 Počet osobních a nákladních vlakových souprav v úseku Mladá Boleslav město, 2016-2019.

7 KATEGORIZACE STANIC

Pro zařazení všech monitorovacích bodů do kategorií byla použita metoda shlukování a zařazení na základě četnosti nákladní a osobní dopravy (viz obr. 7.1). Data byla použita z roku 2016, protože jde o jediná bodová data za celý rok, která byla poskytnuta. Bohužel nelze z úsekových dat získat data bodová, jinak by bylo možné kategorizovat na základě nejnovějších dat z roku 2019. Do kategorizace byly vybrány různé kombinace atributů a následně byla zvolena optimální řešení a ta interpretována. Byly použity jak roční agregace rozdílných atributů, tak měsíční agregace stejných atributů. Výsledné mapové výstupy pro 5 tipů kategorizací byly umístěny do přílohy (příloha 4-8).



Obr. 7.1 Schéma rozdělení kategorizací na základě metody a vstupních dat.

7.1 Roční agregace

7.1.1 A. Procentuální podíl dopravy

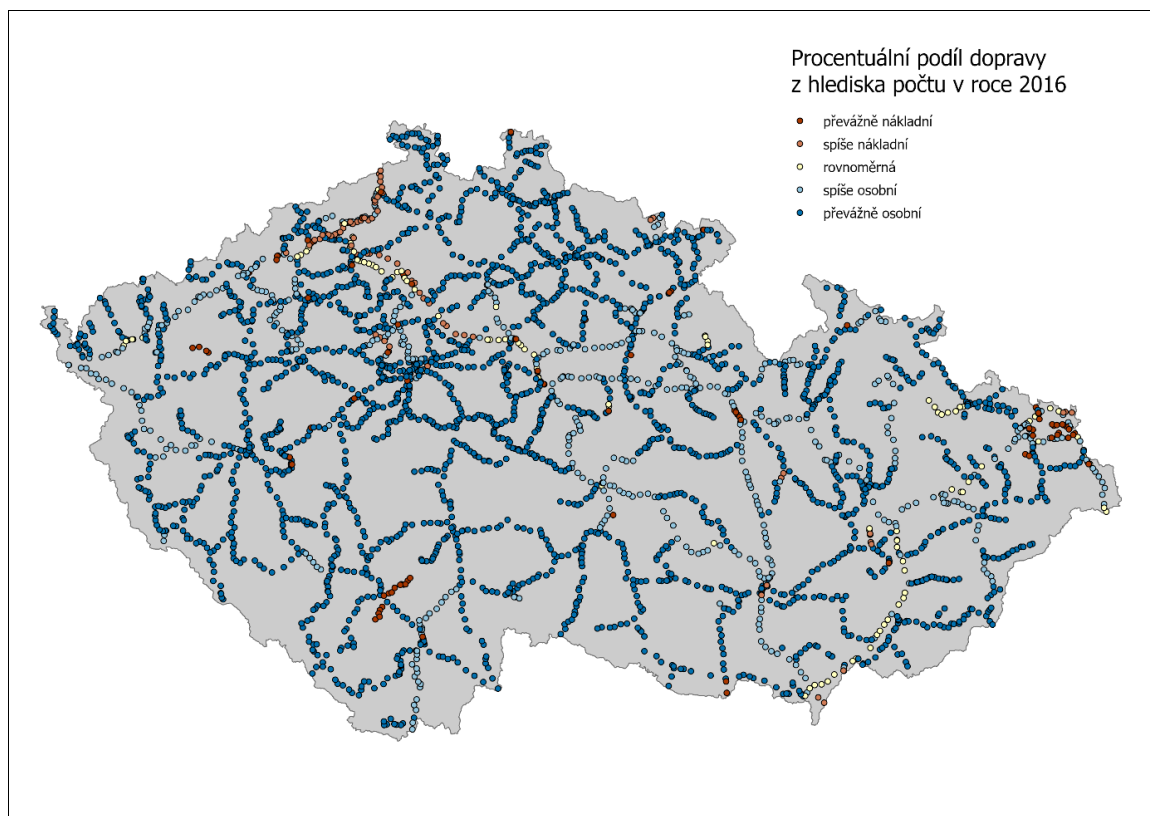
Jako první byla vyhotovena základní kategorizace stanic z hlediska váhy a počtu osobních a nákladních vlakových souprav a jejich zastoupení ve stanicích po celé ČR. Pro toto zpracování nebyly atributy dále zpracovávány, pouze bylo vypočítáno procentuální zastoupení osobní a nákladní dopravy. Byla zvolena obdobná logika jako u tvorby výstupů na liniových datech. Bylo vytvořeno pět kategorií pro jednotlivá zastoupení osobní a nákladní dopravy z hlediska hmotnosti a jejich počtu vždy po 20 %. To znamená, že pokud je zastoupena nákladní doprava hmotností ve stanici 20 % tak osobní doprava je zastoupena 80 % a dále analogicky. Vzniká tak možnost pro tvorbu kategorií na základě tohoto zastoupení. Následující škála byla použita pro hmotnost i počet (tab. 6).

Tab. 6 Kategorie procentuální kategorizace poměru počtů a hmotností osobní a nákladní dopravy

interval [%]	kategorie
80–100	převážně osobní
60–79	spíše osobní
40–59	rovnoměrná doprava
20–39	spíše nákladní
0–19	převážně nákladní

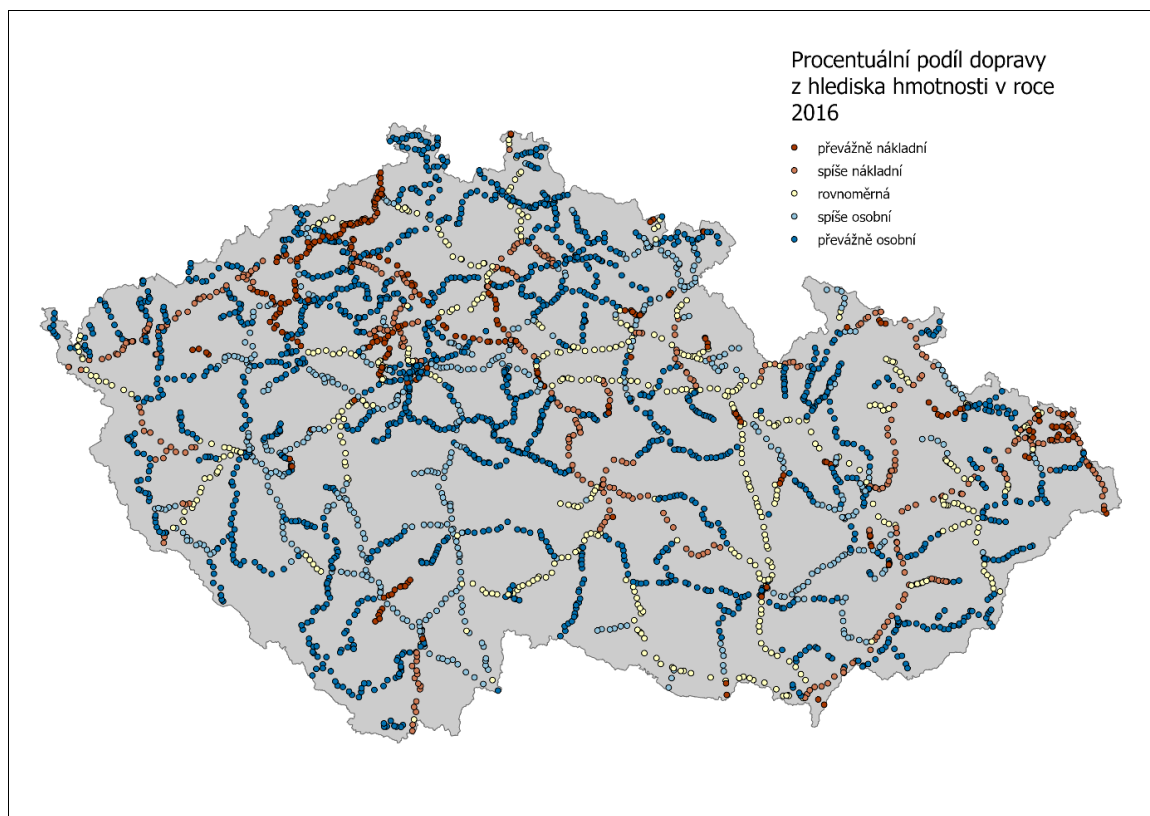
V této kategorizaci se neuvažuje celkový počet nebo hmotnost v dané stanici (monitorovacím bodu), ale pouze relativní procentuální podíl hmotnosti nebo počtu vlakových souprav. Nedává tedy informaci o absolutním zatížení trati, ale pouze přehled o struktuře dopravy z hlediska hmotnosti a počtu.

Aby bylo možné, z hlediska prostoru, lépe interpretovat výsledky kategorizace, byly vytvořeny datové náhledy obou typů kategorizací. Jako první vznikl náhled pro vyjádření poměru dopravy z hlediska počtu osobních a nákladních vlakových souprav za rok 2016. Na tomto náhledu (viz obr. 7.2) je možné vidět převažující kategorii převážně osobní, zobrazenou tmavě modrou barvou. Tato kategorie informuje o tom, kde je z hlediska počtu 80–100 % osobních vlakových souprav vzhledem k nákladním. Další kategorie, spíše osobní je již zastoupena méně, ale z hlediska všech kategorií také poměrně významně. Zde se osobní doprava už více mísí s dopravou nákladní. Vyrovnaná situace v poměru dopravy, kterou reprezentuje kategorie rovnoměrná doprava, se vyskytuje opět v menším množství než předchozí kategorie a převážně na tranzitních koridorech. Zejména pak na 2. tranzitním koridoru z Přerova do Břeclavi. Dalším významným místem výskytu této kategorie je trať z Kolína do Děčína, která je druhou trasou evropského nákladního koridoru RFC 7 (Východní a východo – středomořský koridor), jehož první větev vede přes vytiženou Prahu. Kategorie spíše nákladní je převážně zastoupena na dvou místech, prvním je již zmíněný odklon evropského nákladního koridoru RFC 7 z Kolína do Děčína. Druhý úsek je od Biliny, kde se nachází stejnojmenný hnědohelný důl, až na hranice s Německem do monitorovacího bodu Děčín státní hranice. Oba úseky se nacházejí na evropském nákladním koridoru RFC 7, ale také na evropském nákladním koridoru RFC 8 (Severomořsko-baltský koridor). Poměrně nejvytíženější, počtem nákladní dopravy oproti osobní, je oblast Ostravska a poté několik málo oblastí po celé ČR, které jsou většinou počátkem například soukromé vlečky, nebo vlečkou samotnou. Dále jde o oblasti, kde je zhuštěná nákladní doprava kvůli průmyslu, nebo těžbě dřeva a jeho zpracování. Nemusí jít o tratě s vysokým absolutním počtem nákladních vlakových souprav. Poměrně často na těchto tratích není zavedena pravidelná osobní doprava. Výsledný mapový výstup je umístěn v příloze (příloha 4).



Obr. 7.2 Procentuální podíl dopravy z hlediska počtu vlakových souprav v roce 2016.

Jako druhý vznikl náhled pro vyjádření poměru dopravy z hlediska hmotnosti osobních a nákladních vlakových souprav za rok 2016 (viz. obr. 7.3). V hodnocení této kategorizace, si je nutné uvědomit, že průměrná váha nákladního vlakové soupravy je násobně vyšší než průměrná váha vlakové soupravy osobního. Nelze zde sice hodnotit absolutní čísla, ale na tomto výstupu je dobře vidět, že zde už není tak převažující osobní doprava. Rozložení všech kategorií je zde přibližně rovnoměrné. Většina tranzitních koridorů v ČR je naplněna kategoriemi spíše osobní, rovnoměrná a spíše nákladní. Kategorie převážně osobní se zde nachází zcela výjimečně a kategorie převážně nákladní se vyskytuje na koridorech převážně na Ústecku a Děčínsku. Převážně osobní se dále vyskytuje na místech lokálních tratí nebo vysoké osobní přepravy na tratích, které jsou z hlediska přepravy nákladní bezvýznamné. Prakticky jde říci, že oproti kategorizaci na základě počtu zde logicky vystupuje více do popředí doprava nákladní z důvodu její větší průměrné váhy. Také tato kategorizace přináší lepší pohled na opravdovou strukturu zatížení tratí, protože jde o poměr hmotností. Samozřejmě podle této kategorizace nelze říci, jak moc je trať zatížená, ale lze říci jakou dopravou to je převážně způsobeno.



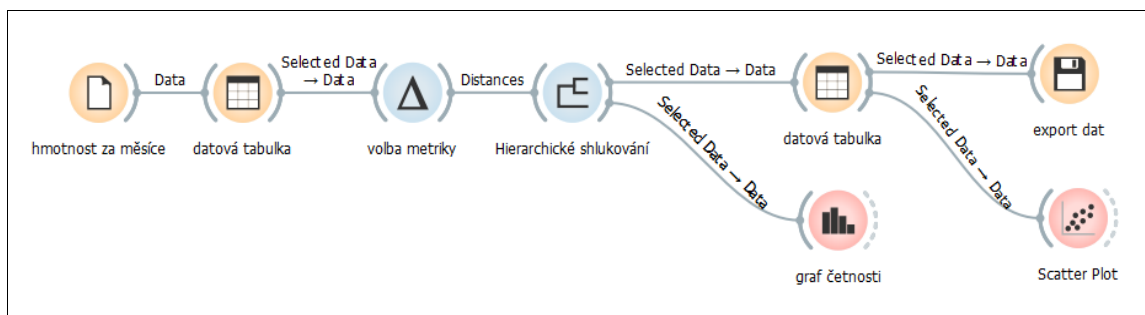
Obr. 7.3 Procentuální podíl dopravy z hlediska hmotnosti vlakových souprav v roce 2016.

7.1.2 B. Shlukování

Dalším způsobem zvoleným v této práci, byla kategorizace provedená formou shlukování. Pro shlukování ročních agregací dat byly zvoleny následující 2 způsoby kombinace atributů:

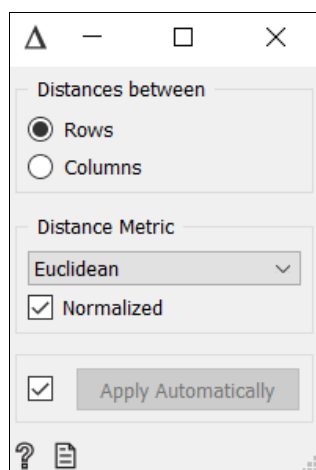
- počet osobních a nákladních vlakových souprav,
- hmotnost osobních a nákladních vlakových souprav.

Dále bylo vyzkoušeno zpracování všech atributů najednou, kde byla navíc použita analýza hlavních komponent. Výsledky této kombinace však nebyly uspokojivé, tak bylo přistoupeno k vyřazení tohoto výběru. Shlukování bylo realizováno v programu Orange, kde byl vyhotoven model (viz obr. 7.4) od nahrání dat po jejich export, pro další použití ve formátu CSV. Konkrétně šlo o import vstupních dat ve formátu CSV, poté volba metriky a normalizace dat a následně volba shlukovací metody a export. Pro sledování výsledků jednotlivých shlukování byly přidány datové tabulky, ve kterých jsou vždy uloženy výsledky předchozího kroku a grafy pro názornou ukázkou.



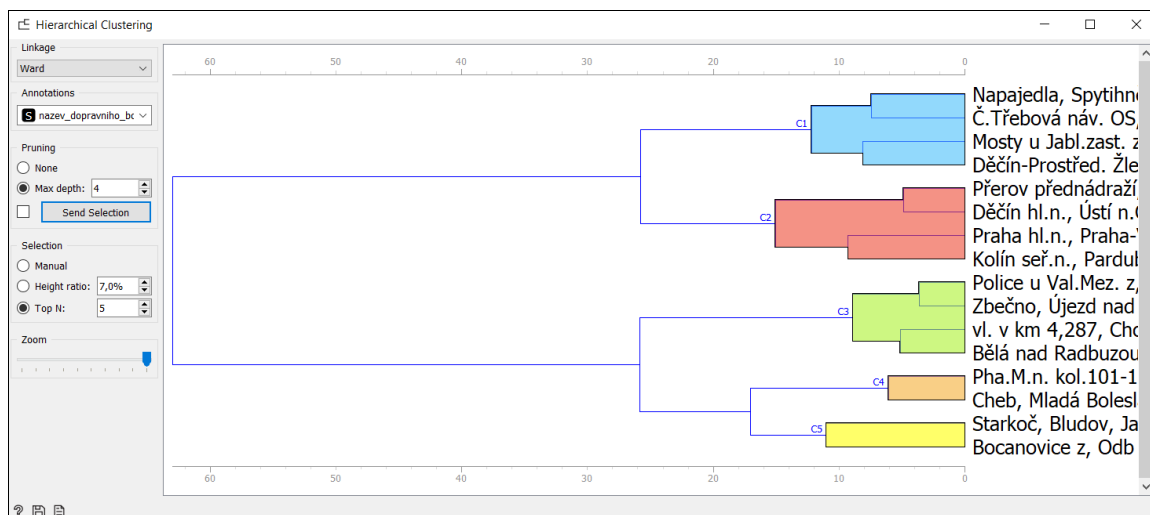
Obr. 7.4 Model hierarchického shlukování v programu Orange.

V rámci nástroje *Distance* byly voleny parametry (viz obr. 7.5), které jsou důležité pro vstup do shlukování. Prvním parametrem je volba, mezi čím chceme počítat vzdálenost. V rámci tohoto nástroje je možnost zvolit, zdali chceme počítat vzdálenost mezi sloupci, nebo řádky. Vzhledem k tomu, že data mají strukturu, kdy každý sloupec znamená jeden atribut a každý řádek jeden záznam, byla zvolena možnost řádky. Dále bylo nutné zvolit, jakou chceme počítat vzdálenost. Zde bylo velké množství různých alternativ, v tomto případě byla vybrána vzdálenost *euklidovská*. Poslední volbou bylo, zda chceme data normalizovat. Pro shlukování je normalizace nutná, takže tato volba byla potvrzena. Výsledkem nástroje *distance* je vzdáleností matice dat.



Obr. 7.5 Parametrizace nástroje *Distance* v programu Orange.

Poté byla provedena parametrizace nástroje hierarchické shlukování (viz obr. 7.6). Nejdříve byla zvolena samotná metoda shlukování. Po předchozím testování jednotlivých metod byla zvolena *Wardova* metoda, protože vykazovala nejsrozumitelnější výsledky. Také jde o nejcitovanější a nejpoužívanější metodu pro shlukování popsanou v odborných článcích a pracích. Dále je možné zvolit, jakým atributem ze zdrojových dat chceme popsat jednotlivé záznamy ve výsledném shlukovacím stromu. Poté je možnost zvolit stupeň větvení shlukovacího stromu. Tímto parametrem se nastavuje, jaký maximální počet rozvětvení může strom obsahovat. Dalším parametrem je pak volba počtu shluků, to je možné třemi způsoby. První způsob je manuální, formou klikání na myš, označování jednotlivých shluků. Dalším způsobem je procentuální označení zleva doprava, kde je umístěna pomyslná dělicí čára (hladina), od níž napravo jdou jednotlivé shluky. Poslední možností je manuální vepsání počtu shluků. Poslední možnost byla zvolena v této práci s hodnotou pět.



Obr. 7.6 Parametrizace a dendrogram hierarchického shlukování v programu Orange.

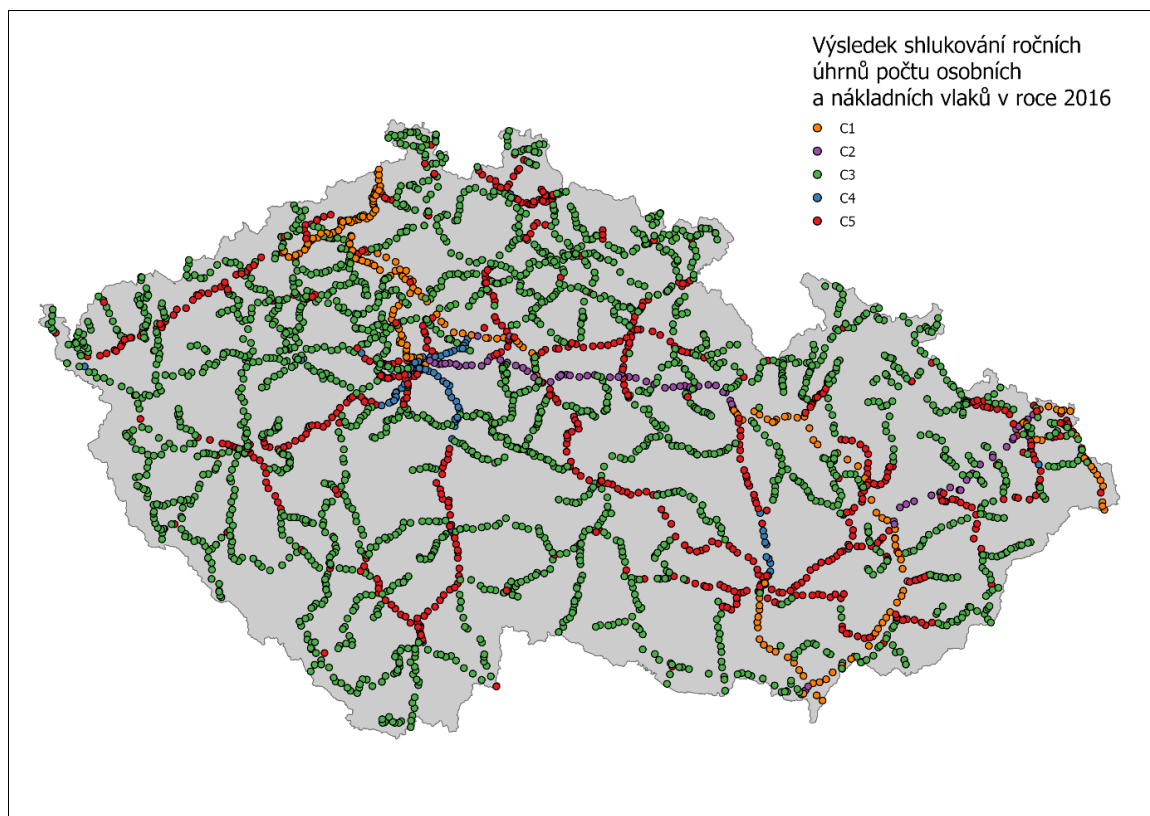
Stejně jako u předchozího způsobu kategorizace na základě poměru osobní a nákladní dopravy, byly vyhotoveny náhledy na výslednou kategorizaci. Jak je popsáno výše, výsledné kategorizace formou shlukování se dělí do pěti kategorií. Pro shluky od jedné do 5 byla zvolena stejná barevná škála pro všechny typy výstupů ze shlukování. Tyto výstupy nejsou porovnatelné, tudíž není přímé srovnání mezi shlukem C1 v prvním náhledu a shlukem C1 v druhém náhledu. Do jisté míry jsou si některé stejně označené shluky podobné, ale to vychází zejména z toho, že jsou vyhotoveny z částečně korelujících parametrů, a tudíž podávají částečně podobné výsledky. Cílem pro interpretaci výsledků, bylo pojmenování jednotlivých shluků dle podstaty a povahy shluku. Toho bylo docíleno pomocí:

- dendrogramu shlukování (popis podobnosti mezi shluky na základě polohy větvení),
- náhledu (jaké typy tratí jsou dotčeny jednotlivými kategoriemi),
- grafem průměrů vstupních dat pro jednotlivé kategorie (průměrné hodnoty pro danou kategorii v absolutním a procentuální vyjádření), grafem vstupních dat zařazených do jednotlivých shluků (potvrzení výše zmíněných průměrů).

Tab. 7 Intervaly počtu a hmotnosti osobních a nákladních vlakových souprav pro prvotní pojmenování shluků

kategorie	počet vlakových souprav	hmotnost vlakových souprav (tis. t)
velmi vysoké	55 000 a více	30 000 a více
vysoké	35 000 – 54 999	25 000 – 29 999
střední	20 000 – 34 999	20 000 – 24 999
malé	10 000 – 19 999	10 000 – 19 999
velmi malé	2 000 – 9 999	5 000 – 9 999
minimální	méně než 2 000	méně než 5 000

Šlo o pojmenování z hlediska typického reprezentanta shluku, tudíž byl rozptýl jednotlivých kategorií zobrazen průměry, jakožto těžišti jednotlivých shluků. Průměrné hodnoty byly zařazeny do kategorií (viz tab. 7) pro lepší pojmenování a viditelnost rozdílů.



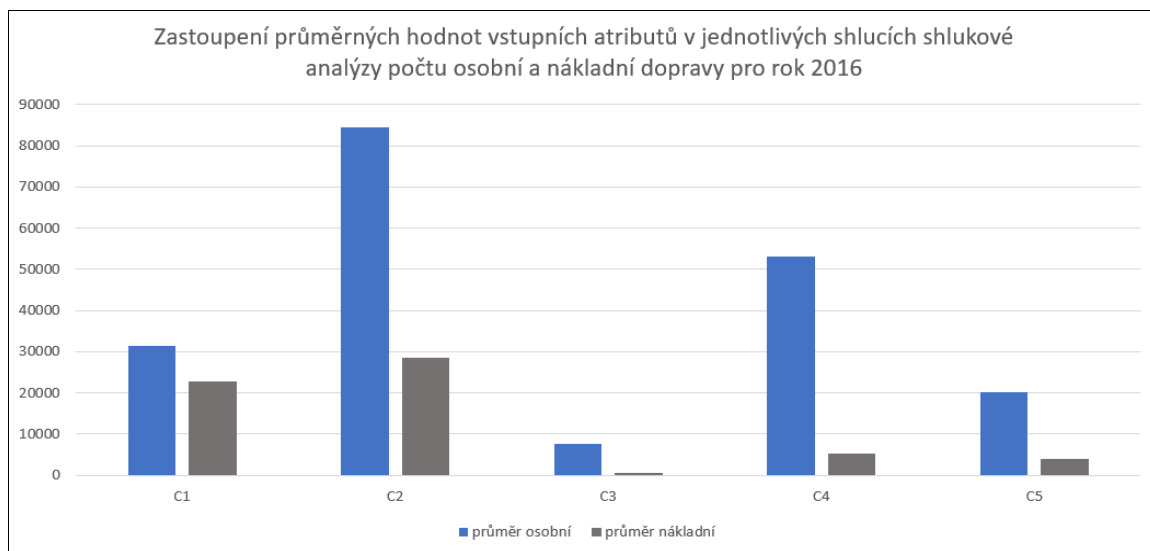
Obr. 7.7 Výsledek shlukování ročních úhrnů počtu osobních a nákladních vlaků v roce 2016.

B1. Vyhodnocení shlukování ročních úhrnů počtu osobních a nákladních vlakových souprav

První vyhodnocovaným typem shlukování ročních dat byl ten, na jehož vstupu byly počty osobních a nákladních vlakových souprav (viz obr. 7.7). Vlastnosti nově vzniklých shluků z atributů počtů osobní a nákladní dopravy, definovaných pomocí dendrogramu, náhledu a grafu průměrů vstupních dat pro jednotlivé kategorie (viz obr. 7.8):

- C1
 - nejvíce podobné se shlukem C2, méně s ostatními shluky,
 - převážně monitorovací body na tranzitním koridoru a v menším množství monitorovací body, které jsou na trasách evropských nákladních koridorů,
 - mírná převaha osobní dopravy (cca 32 000 vlakových souprav – 58 %) nad nákladní (cca 23 000 vlakových souprav – 42 %), oba typy dopravy ve středních hodnotách průměrných počtů vlakových souprav.

- C2
 - nejvíce podobná se s hlukem C1, méně s ostatními shluky,
 - pouze na hlavním koridoru v úsecích Praha – Česká Třebová a Olomouc – Ostrava,
 - násobná převaha osobní dopravy (cca 84 000 vlakových souprav – 75 %) nad nákladní (cca 29 000 vlakových souprav – 25 %), osobní ve velmi vysokých průměrných hodnotách počtu vlakových souprav a nákladní ve středních.
- C3
 - nejvíce podobná se shluky C4 a C5, méně s ostatními,
 - převážně regionální tratě, nebo tratě vyššího významu, avšak s malým objemem dopravy,
 - naprostá převaha osobní dopravy (cca 7 600 vlakových souprav – 97 %) nad nákladní (cca 600 vlakových souprav – 3 %), osobní doprava ve velmi malých průměrných hodnotách počtu vlakových souprav a nákladní v minimálních.
- C4
 - nejvíce podobná se shlukem C5, méně s C4, a ještě méně s ostatními,
 - několik úseků v okolí Prahy v dosahu příměstské dopravy,
 - naprostá převaha osobní dopravy (cca 53 000 vlakových souprav – 91 %) nad nákladní (cca 5 300 vlakových souprav – 9 %), osobní doprava ve vysokých průměrných hodnotách počtu vlakových souprav a nákladní ve velmi malých.
- C5
 - nejvíce podobná se shlukem C4, méně s C5, a ještě méně s ostatními,
 - tranzitní koridory (s výjimkou kat. C1 a C2), dálkové evropské koridory a významné železniční tahy,
 - převaha osobní dopravy (cca 20 000 vlakových souprav – 83 %) nad nákladní (cca 4 000 vlakových souprav – 17 %), osobní doprava v malých průměrných hodnotách počtu vlakových souprav a nákladní ve velmi malých.



Obr. 7.8 Průměrné hodnoty vstupních atributů v jednotlivých shlucích počtu osobní a nákladní dopravy pro rok 2016.

Na základě shlukové analýzy a následného vyhodnocení došlo na pojmenování jednotlivých shluků do podoby, jež odpovídá jeho vlastnostem. Uvažovány byly typy tratí, poměr typu dopravy a průměrný počet vlakových souprav zařazený do kategorií (viz tab. 8).

Tab. 8 Intervaly průměrného počtu vlakových souprav celkem pro finální pojmenování shluků

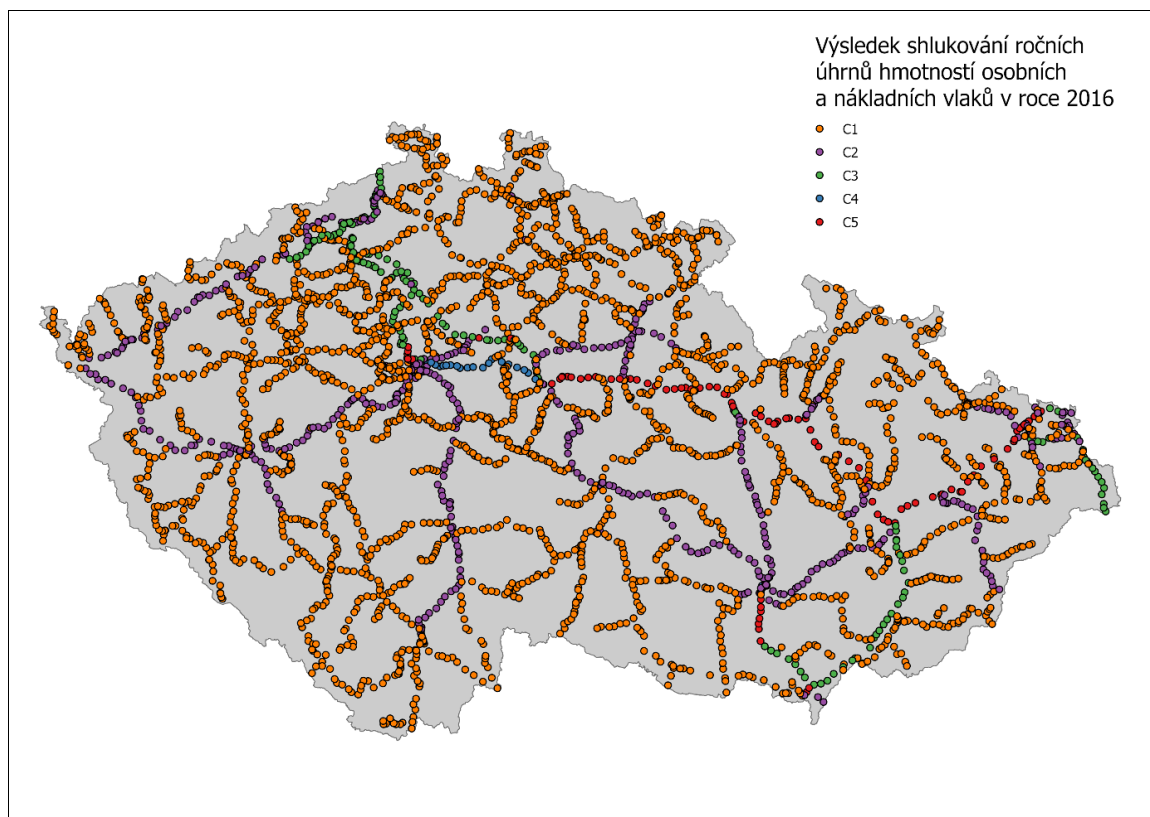
typ	interval počtu vlakových souprav
velmi vysoký	80 000 a více
vysoký	60 000 – 79 999
průměrný	40 000 – 59 999
malý	20 000 – 39 999
velmi malý	méně než 20 000

Na základě shlukování počtu osobních a nákladních vlakových souprav vznikly tyto kategorie, které jsou seřazeny z hlediska průměrného počtu všech vlakových souprav dané kategorie sestupně.

- C2: Stanice na tranzitních koridorech s vysokou převahou osobní dopravy a velmi vysokým počtem vlakových souprav
- C4: Stanice s příměstskou dopravou s dominantní osobní dopravou a průměrným počtem vlakových souprav
- C1: Stanice na nákladních a osobních koridorech s mírnou převahou osobní dopravy a průměrným počtem vlakových souprav
- C5: Stanice na významných tratích s dominantní osobní dopravou s malým počtem vlakových souprav
- C3: Stanice na regionálních a nadregionálních tratích s dominantní převahou osobní dopravy s velmi malým počtem vlakových souprav

B2. Vyhodnocení shlukování ročních úhrnů hmotnosti osobních a nákladních vlakových souprav

Druhým vyhodnocovaným typem shlukování ročních dat byl ten, na jehož vstupu byly počty osobních a nákladních vlakových souprav. Interpretační postup byl dodržen stejný jako u předchozího případu se shlukováním počtu osobních a nákladních vlakových souprav.



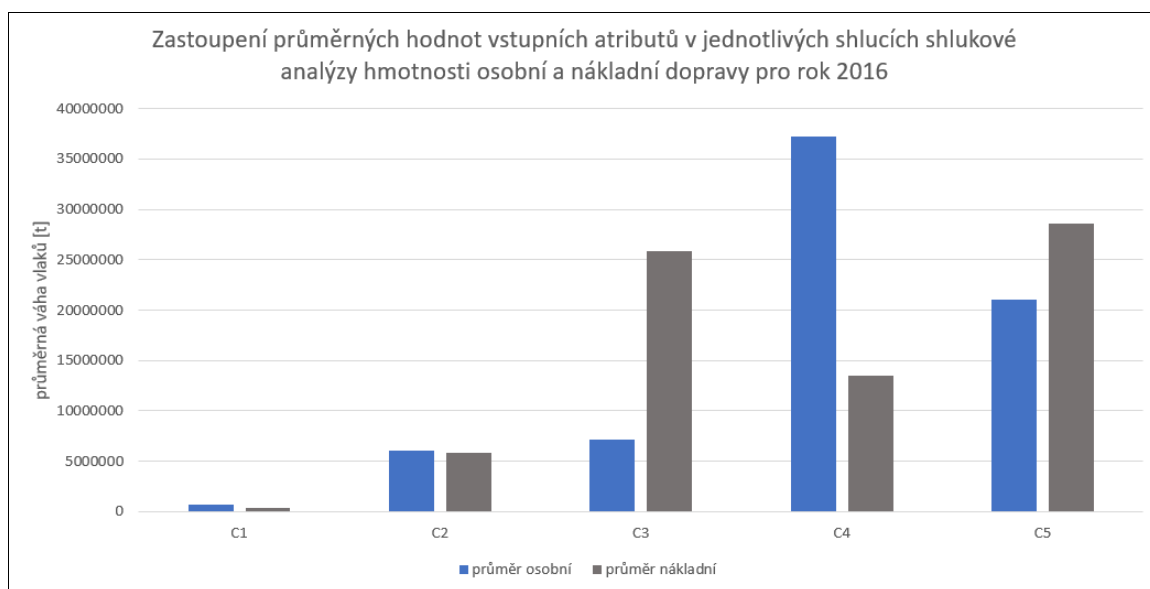
Obr. 7.9 Výsledek shlukování ročních úhrnů hmotnosti osobních a nákladních vlaků v roce 2016.

Vlastnosti nově vzniklých shluků z atributu hmotnosti osobní a nákladní dopravy (viz obr. 7.9), definovaných pomocí dendrogramu, náhledu a grafu průměrů vstupních dat pro jednotlivé kategorie (viz obr. 7.10):

- C1
 - nejvíce podobné se shlukem C2, méně s ostatními shluky,
 - regionální tratě a některé tratě vyššího významu s malým objemem dopravy,
 - převaha osobní dopravy (cca 6 900 000 t – 63 %) nad nákladní (cca 4 000 000 t – 37 %), osobní doprava ve velmi malých hodnotách průměrné hmotnosti vlakových souprav a nákladní v minimální.
- C2
 - Nejvíce podobná se s hlukem C1, méně s ostatními shluky
 - Části tranzitních koridorů a evropských nákladních koridorů po celé ČR, oblast Podkrušnohoří a trať Hranice na Moravě – Horní Lideč
 - Téměř vyrovnaný stav osobní (cca 6 100 000 t – 51 %) a nákladní dopravy (cca 5 900 000 t – 49 %), oba typy dopravy ve velmi malých hodnotách průměrné hmotnosti vlakových souprav
- C3
 - Nejvíce podobná se shluky C4 a C5, méně s ostatními
 - Část tranzitních koridorů a evropských nákladních koridorů severně a severovýchodně od Prahy směrem na hraniční přechod se SRN u Děčína a také trať Mostu – Ústí nad Labem
 - Převaha nákladní dopravy (cca 25 900 000 t – 78 %) nad osobní (cca 7 200 000 t – 22 %), osobní doprava ve velmi malých hodnotách průměrné hmotnosti vlakových souprav a nákladní ve vysokých
- C4
 - Nejvíce podobná se shlukem C5, méně s C4, a ještě méně s ostatními
 - Hlavní tranzitní koridor v úseku Praha – Kolín
 - Převaha osobní dopravy (cca 37 200 000 t – 73 %) nad nákladní (cca 13 500 000 t – 27 %), osobní doprava ve velmi vysokých hodnotách průměrné hmotnosti vlakových souprav a nákladní v malých

C5

- Nejvíce podobná se shlukem C4, méně s C5, a ještě méně s ostatními
- Hlavní tranzitní koridor v úseku Kolín – Ostrava a poté malé menší fragmenty tranzitních koridorů u Brna a Prahy
- Převaha nákladní dopravy (cca 28 600 000 t – 58 %) nad osobní (cca 21 100 000 t – 42 %), osobní doprava ve středních hodnotách průměrné hmotnosti vlakových souprav a nákladní v malých



Obr. 7.10 Průměrné hodnoty vstupních atributů v jednotlivých shlucích hmotnosti osobní a nákladní dopravy pro rok 2016.

Také na základě výstupů z druhé varianty shlukování ročních dat bylo vytvořeno vyhodnocení jednotlivých shluků. Stejně jako v předchozím případě byly použity typy tratí a poměry typu dopravy. Průměrný počet všech vlakových souprav byl nahrazen průměrnou hmotností všech vlakových souprav v dané kategorii (viz tab. 9).

Tab. 9 Intervaly průměrné hmotnosti vlakových souprav celkem pro finální pojmenování shluků

typ	interval hmotnosti všech vlakových souprav [tis. t]
velmi vysoký	40 000 a více
vysoký	30 000 – 39 999
průměrný	20 000 – 29 999
malý	10 000 – 19 999
velmi malý	méně než 10 000

Na základě shlukování hmotnosti osobních a nákladních vlakových souprav vznikly tyto kategorie, které jsou seřazeny z hlediska průměrné hmotnosti všech vlakových souprav dané kategorie sestupně.

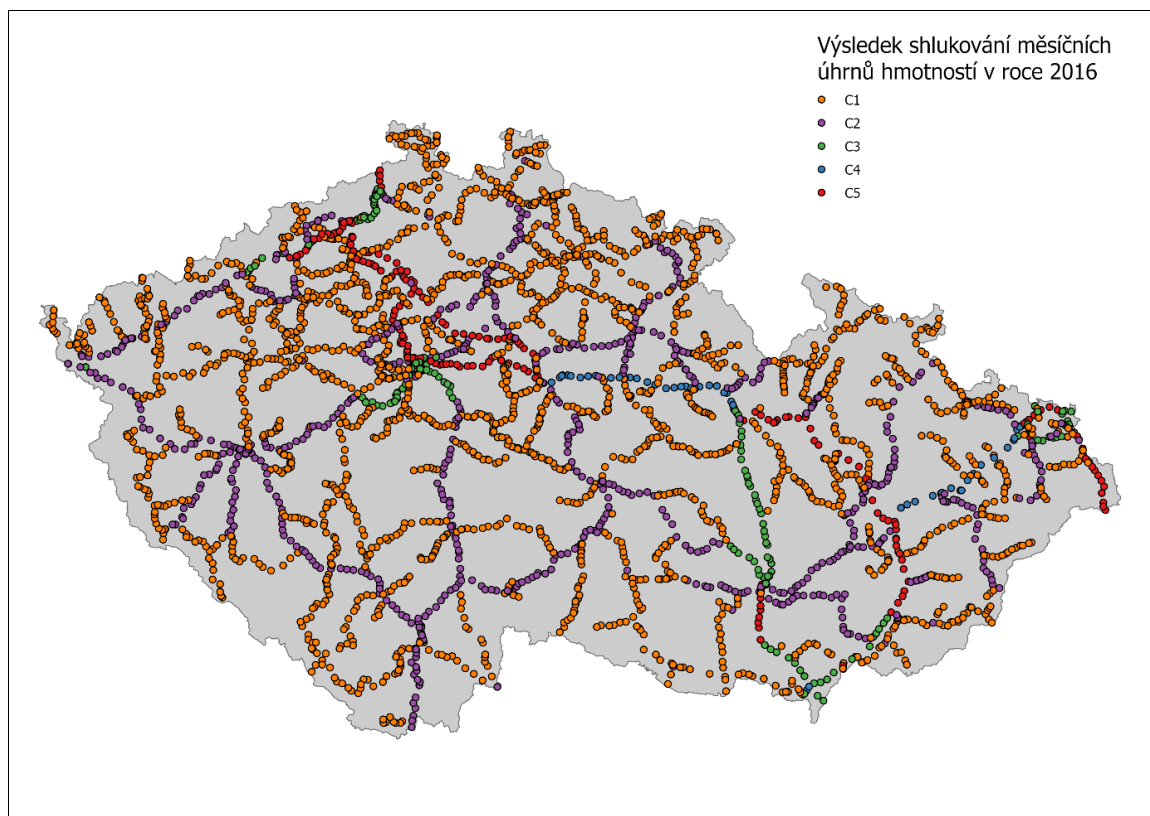
- C4: Stanice na tranzitních koridorech s dominantní osobní dopravou a velmi vysokou celkovou hmotností vlakových souprav
- C5: Stanice na tranzitních koridorech s převahou nákladní dopravy a velmi vysokou celkovou hmotností vlakových souprav
- C3: Stanice na osobních a nákladních koridorech s dominantní nákladní dopravou a vysokou celkovou hmotností vlakových souprav
- C2: Stanice na osobních a nákladních koridorech a významných tratích s nepatrnou převahou osobní dopravy a malou celkovou hmotností vlakových souprav
- C1: Stanice na regionálních a nadregionálních tratích s převahou osobní dopravy a velmi malou celkovou hmotností vlakových souprav

7.2 Měsíční agregace

7.2.1 C. Shlukování

Na základě měsíčních agregací dat byla také vytvořena kategorizace formou shlukování. Konkrétně byly, z důvodu popsanych v úvodu této kapitoly, vybrány měsíční data za rok 2016. Jde tedy o 12 hodnot jednoho atributu. Protože je v této práci řešeno především zatížení železniční infrastruktury, které reprezentuje více než cokoli jiného ze vstupních dat právě hmotnost, byl vybrán právě tento atribut. Výsledná kategorizace by měla přinést nadhled nejen nad celkovým rozdělením na základě absolutní hmotnosti, ale také nad podobností v ročním průběhu těchto hodnot. Obdobně jako v případě agregací za rok, byl použit pro shlukování program Orange, ve kterém byl vybudován model. Model je totožný jako pro shlukování ročních úhrnů, ale vzhledem k tomu, že jednotka všech atributů je stejná a byly použity hodnoty pouze jednoho atributu rozdělené do 12 měsíců, tak nebylo potřebné odstraňovat vliv rozdílných jednotek.

Nastavení všech nástrojů v modelu proběhlo téměř totožně jako u shlukování ročních dat, s tím rozdílem, že vstupní data byla v úhrnech za měsíce pouze pro jeden atribut. Hodnocení proběhlo obdobně jako u ročních dat. Použit byl dendrogram a náhled pro popsání vlastností jednotlivých shluků. Vzhledem k tomu, že je použit pouze jeden atribut pro 12měsíčních hodnot hmotnosti pro všechny vlakové soupravy, tak nebylo použito porovnání jednotlivých druhů dopravy formou grafu, jako v přechozím případě. Graf byl zhotoven pro průměrné hodnoty hmotnosti pro každou kategorii za jednotlivé měsíce (viz. obr. 7.12). Jde pouze o přibližný pohled na průměrné hodnoty, kterých nabývají jednotlivé kategorie. Průměr poměrně výrazně vyhlazuje variabilitu této 12měsíční časové řady. Graf nám tak spíše ukazuje průměrnou hladinu absolutních hodnot hmotnosti. Vzhledem k tomu, že vzniklé časové záznamy průměrných hodnot jsou poměrně ekvivalentní, bude přistoupeno k vytvoření kategorie pro každý shluk z hlediska jeho průměrné hodnoty.



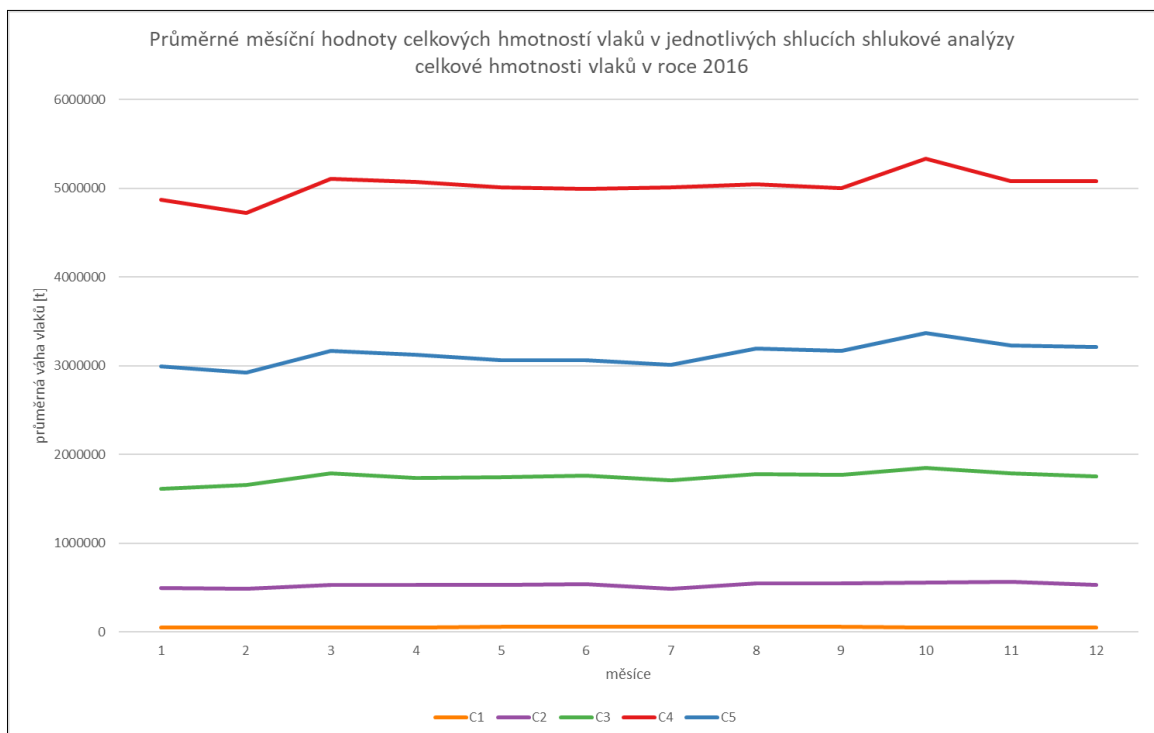
Obr. 7.11 Výsledek shlukování měsíčních úhrnů hmotnosti vlakových souprav roce 2016.

C1. Vyhodnocení shlukování měsíčních úhrnů hmotností osobních a nákladních vlakových souprav

Vlastnosti nově vzniklých shluků z atributu hmotnosti vlakových souprav (viz obr. 7.11) byly definovány pomocí výstupů získaných jako výsledek ze shlukování.

- C1
 - nejvíce podobné se shlukem C2, méně s ostatními shluky,
 - regionální tratě a některé tratě vyššího významu s malým objemem dopravy,
 - velmi malá průměrná hmotnost vlakových souprav za měsíc, přibližně 55 000 t.
- C2
 - nejvíce podobná se shlukem C1, méně s ostatními shluky,
 - koridorové úseky Praha – České Budějovice a Praha – Plzeň a poté některé vytíženější tratě po celé ČR,
 - malá průměrná hmotnosti vlakových souprav za měsíc, přibližně 531 000 t.

- C3
 - nejvíce podobná se shluky C4 a C5, méně s ostatními,
 - úseky na koridorech a významných tratích, například Česká Třebová – Brno, okolí Břeclavi, Prahy, nebo Podkrušnohoří,
 - střední průměrná hmotnost vlakových souprav za měsíc, přibližně 1 746 000 t.
- C4
 - nejvíce podobná se shlukem C5, méně s C4, a ještě méně s ostatními,
 - pouze koridorové úseky Kolín – Česká Třebová a Přerov – Ostrava,
 - velmi velká průměrná hmotnost vlakových souprav za měsíc, přibližně 5 027 000 t.
- C5
 - nejvíce podobná se shlukem C4, méně s C5, a ještě méně s ostatními,
 - pouze koridorové úseky a úseky Evropského nákladního koridoru, Kolín –Praha – Ústí nad Labem, Kolín – Ústí nad Labem,
 - velká průměrná hmotnost vlakových souprav za měsíc, přibližně 3 127 000 t.

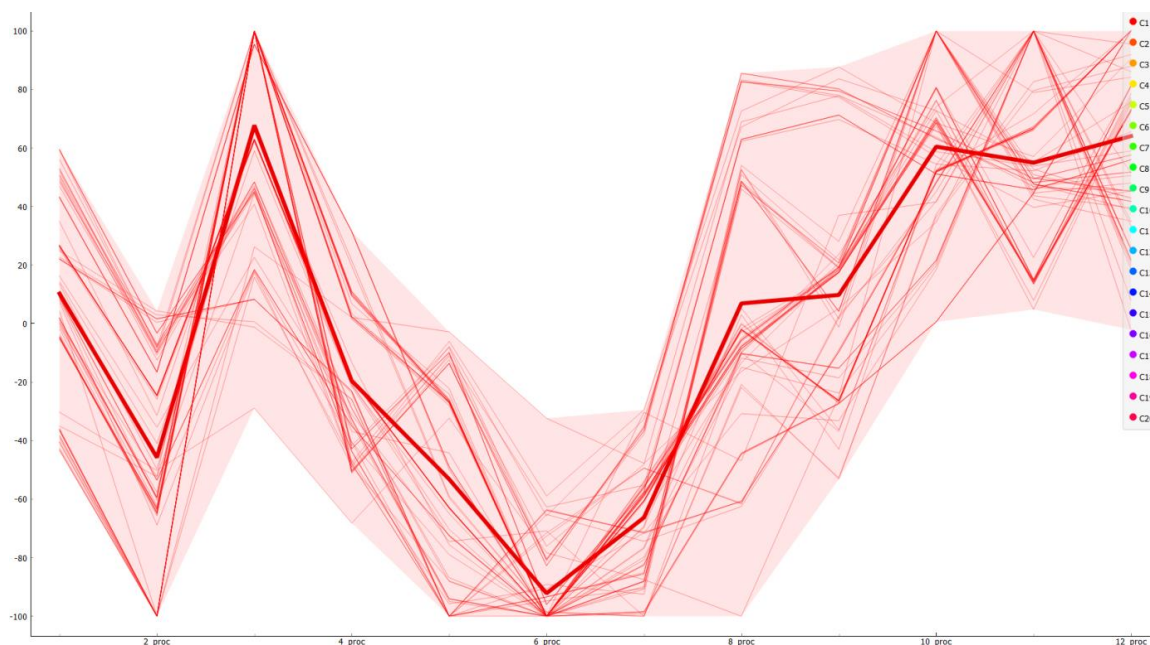


Obr. 7.12 Průměrné hodnoty vstupních atributů v jednotlivých shlucích hmotnosti vlakových souprav pro jednotlivé měsíce roku 2016.

Nakonec bylo přistoupáno také k pojmenování shluků z poslední shlukové analýzy celkových měsíčních hmotností vlakových souprav. Pojmenování proběhlo podle dříve použitého postupu, který byl použit pro pojmenování shluků z ročních dat.

- C1: Stanice na regionálních a méně vytižených nadregionálních tratích s velmi malou hmotností vlakových souprav
- C2: Stanice na méně vytižených koridorech a vytižených nadregionálních tratích s malou hmotností vlakových souprav
- C3: Stanice na vytižených tranzitních koridorech a nadregionálních tratích s průměrnou hmotností vlakových souprav
- C5: Stanice na velmi vytižených nákladních a osobních koridorech s vysokou hmotností vlakových souprav
- C4: Stanice na velmi vytižených tranzitních koridorech s velmi vysokou hmotností vlakových souprav

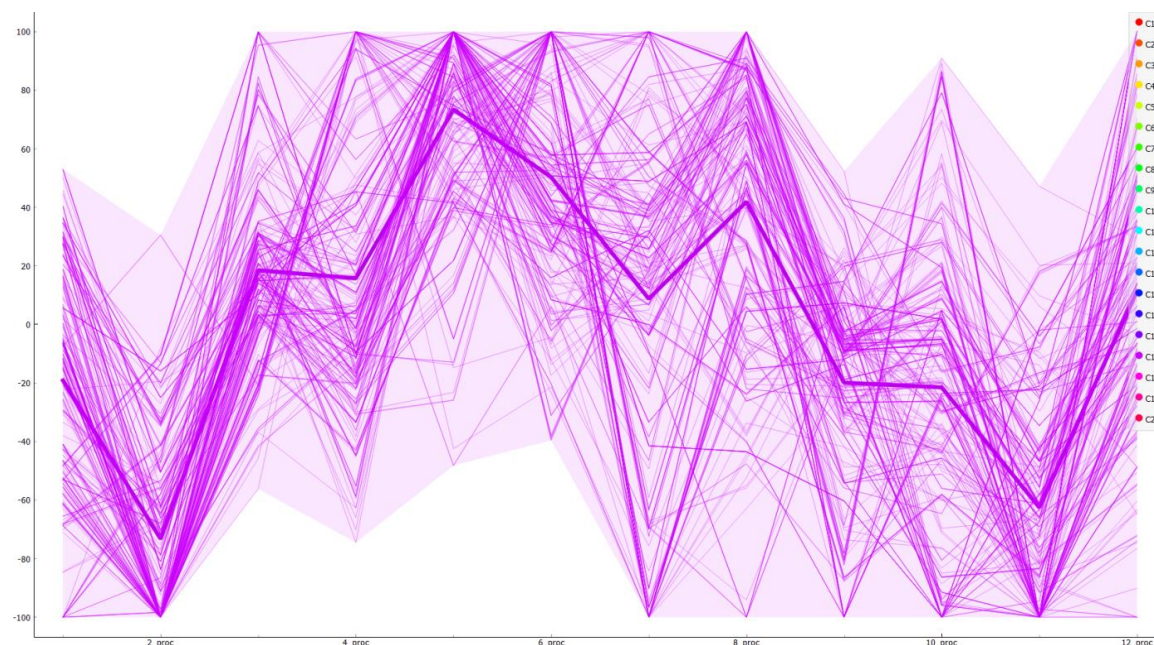
V rámci shlukování dat měsíčních hmotností bylo přistoupáno, z důvodu nespokojenosti s výsledky této analýzy, k úpravě vstupních parametrů na relativní procentuální výchylky od průměrné hodnoty. Upravil jsem měsíční hodnoty tak, aby nabývaly $\pm 100\%$ od průměrné hodnoty konkrétní řady. Takováto reprezentace dat by měla už spolehlivě reprezentovat pouze křivku ročního průběhu po měsících, bez ohledu na absolutní hodnoty. Takto upravené hodnoty hmotností vlakových souprav byly použity jako vstup do shlukování. Po testování různých kombinací typu metriky a shlukovací metody bylo dospěno k závěru, že na výsledek mají jen velmi malý vliv. Největší vliv měl počet výsledných shluků. Teoreticky byl vyzkoušen nejvyšší počet shluků 20. Méně početné shluky tvořili poměrně kompaktní skupiny s malým rozptylem v rámci měsíců (viz obr. 7.13).



Obr. 7.13 Méně početný shluk při shlukování relativní dat hmotnosti vlakových souprav o počtu shluků 20.

Více početné shluky (viz obr. 7.14) už měli rozsah výchylek v jeden měsíc až $\pm 100\%$, což je maximální možný rozsah. Tento jev se v závislosti na shluku opakoval i vícekrát

v rámci jednoho roku. Bohužel ani volba metody, metriky a počtu výsledných shluků nezajistily předpokládané výsledky, proto nebyly do finálních kategorizací zahrnuty, zejména pak proto, že v rámci různých kategorií dosahovali rozdílné kvality výsledků. Pozitivním výsledkem pak může být definování několika shluků z celkového počtu jako často se opakujících ročních průběhu (viz obr. 7.13).



Obr. 7.14 Více početný shluk při shlukování relativní dat hmotnosti vlakových souprav o počtu shluků 20.

Na závěr, tak aby bylo vyzkoušeno co nejvíce metod shlukování, které jsou dostupné pro tato data, byl proveden test v nástroji SimRoad 0.9, dostupném na <http://eyetracking.upol.cz/simroad/>. Bohužel nebylo možné použít nástroj pro kompletní dataset kvůli jeho velikosti. S částečnými soubory nástroj fungoval, avšak v tomto konkrétním případě se zdál být nepoužitelným.

7.3 Hodnocení

V rámci kategorizace dat o železniční infrastruktuře vzniklo několik různých přístupů. Tyto kategorizace pomohou lépe porozumět datům o železniční infrastruktuře poskytnutým SŽ ČR. Na jejich základě je možné usuzovat skutečnosti o struktuře, zatížení a prostorovém rozložení počtu, hmotnosti a druhu dopravy na železniční síti ČR v roce 2016.

Prvním přístupem byl přepočítání počtů a hmotností osobních a nákladních vlakových souprav v ročních úhrnech na procentuální hodnoty. Tento jednoduchý postup dal k dispozici jednoduchý a rychlý náhled na situaci rozložení osobní a nákladní dopravy v ČR. Na základě této kategorizace bylo tedy možné definovat monitorovací bod na základě procentuálního rozložení typu dopravy dle počtu nebo hmotnosti vlakových souprav. Základním rysem, který je možné sledovat na těchto dvou kategorizacích je rozdílná průměrná váha osobních a nákladních vlakových souprav. Dále je zajímavé sledovat extrémy, které informují, kde se nachází převážně osobní a kde převážně nákladní doprava. V neposlední řadě je při kombinaci obou výstupů zřetelně vidět chování nákladní dopravy, která je kromě Ostravska situována převážně na oblast tratí Praha – Ústí nad Labem – Děčín, Most – Ústí nad Labem a Kolín – Ústí nad Labem –

Děčín. To ukazuje, že primární nákladní doprava je vedena po severozápadní části Východního a východo-středomořského koridoru (RFC 7) a Severomořsko-baltského koridoru (RFC 8).

Druhým přístupem bylo použití metody shlukování na datech o hmotnosti a počtu osobních a nákladních vlakových souprav v ročních úhrnech. Pro realizaci této metody byl použit program Orange, který podporuje hierarchické shlukování. Byla použita *euklidovská vzdálenost*, *Wardova metoda* a konečné číslo pět výsledných shluků. Na základě výstupů, které vznikly při a na základě shlukování, byly pak tyto shluky pojmenovány, dle jejich vlastností. Pokud byla použita slovní označení počtu nebo intenzity v případě hmotnosti a počtu vlakových souprav jsou pak v tabulce kvantifikována přesnými čísly nebo intervaly. Tato kategorizace už udává informace o absolutních číslech (hmotnosti nebo počtu vlakových souprav), ale dále i o rozložení v dané kategorii nebo o typu trati, na které se nachází. Na základě shlukování vznikly kategorie, jejichž reprezentanti jsou si podobnější se zástupci své kategorie než se zástupci kategorie jiné. Prostřednictvím vypočítaných průměrů pro jednotlivé kategorie vznikl pohled na typického reprezentanta shluku. U obou kategorizací je možné si všimnout některých zajímavých skutečností. Zatímco některé shluky zaujímají široká spektra různých tratí, některé jsou úzce specifikované, a tak vznikají určité vzory chování ukazatelů. Mezi ně jsou počítány počty a hmotnosti vlakových souprav nebo rozložení osobní a nákladní dopravy. Konkrétně lze například v obou případech vidět trať Praha – Ústí nad Labem – Děčín a Kolín – Ústí nad Labem – Děčín kde probíhá dominantní nákladní doprava. Je vidět také rozdílné chování dopravy na tranzitních koridorech, a to nejen v rámci různých koridorů, ale i v rámci jednoho. Je také zřejmá vysoká intenzita dopravy v Podkrušnohoří na trati z mířící z Chebu do Ústí nad Labem, která vede mimo tranzitní koridory a evropské nákladní koridory.

Poslední přístupem, který byl v této práci použit pro kategorizaci, je shlukování měsíčních agregací dat celkové hmotnosti souprav. Pro realizaci metody shlukování byl zvolen stejný postup jako u shlukování ročních agregací. Předpokladem bylo, že kromě samotného atributu hmotnosti pro jednotlivé měsíce, bude shlukování brát v úvahu i relativní průběh 12měsíční křivky, realita byla však jiná. Prvním problémem je, že shlukování bere v úvahu pouze absolutní vzdálenosti. Pokud bude křivka podobného průběhu, ale v diametrálně odlišných absolutních hodnotách, nebude zařazena do stejné kategorie. Ve výsledném náhledu je vidět, že se kategorizace částečně podobá shlukování ročních dat hmotnosti a počtu. Pokud například porovnáme výsledky s kategorizací vzniklou shlukováním dat o hmotnosti za rok, tak kategorie C1 a C2 jsou tematicky podobné, pouze se kategorie C2 rozrostla na některé monitorovací body z kategorie C1. Dalším problémem je množství jednotlivých stanic a počet shluků. V rámci těchto dat vzniká velké množství různých křivek v různých absolutních hodnotách a dle mého názoru je není možné zařadit do, tak malého množství shluků. Pro odstranění vlivu absolutních hodnot na měsíční průběhy, bylo testováno řešení s relativními daty, které však přineslo pouze částečný úspěch z hlediska zařazení do jednotlivých kategorií. I přes použití velkého množství 20 shluků nebyl výsledek pro všechny shluky natolik přesvědčující, aby byl tento postup zařazen do výstupů. Lze tak pouze na základě výsledků shlukování relativních dat pojmenovat některé shluky, které přinesli dostatečně korelující výsledky jednotlivých entit ve shluku.

Dá se předpokládat, že by se výsledek zlepšoval, pokud bych použili relativní data procentuálních výchylek v každém měsíci a více kategorií pro shlukování.

8 VÝSLEDKY

Výsledky této práce mají 3 roviny: datovou, analytickou a interpretační. První část práce byla zaměřena na získání takové podoby z poskytnutých dat, aby byla použitelná v dalších analýzách a zpracovatelné i nad rámec této diplomové práce v GIS a datamining softwarech. Dále pak na získání výsledných datových sad s výslednými kategorizacemi železničních stanic. Druhá část byla zaměřena na provedení analýzy časových řad a kategorizace tak, aby byly zohledněny dostupné metody a postupy. V poslední rovině byl kladen důraz na vhodnou interpretaci zmíněných analýz a kategorizací s ohledem na geografickou polohu železniční sítě v ČR a získané znalosti z problematiky železniční provozu.

Data

Poskytnutá data nebyla v prvotní formě příliš přehledná a vhodná pro další zpracování. Z hlediska potřeby dále data používat v rámci této diplomové práce bylo přistoupeno k jejich úpravě tak, aby byla vhodná pro vybrané analýzy. Pro kategorizace vznikly tabulky ve formátu XLSX s ročními úhrny, poskytnutých dat za rok 2016 s polohovou informací monitorovacího bodu. Souběžně s tabulkami vznikly prostorové třídy prvků, na základě opravených GPS souřadnic, ve formátu Esri geodatabase se stejnými informacemi použitelné pro prostorové analýzy. Pro analýzu časových řad byla vstupní data ponechána v úhrnech za měsíce, ale byly upraveny tak, aby jednotlivé úseky obsahovali informace o počtu vlakových souprav za jednotlivé měsíce let 2016, 2017, 2018 a 2019. Na základě seznámení se s daty vznikl z úsekových dat zjednodušený liniový model železniční sítě, ke kterému mohou být jednoduše připojeny tabulky zpracované pro analýzu časových řad a vznikne tak prostorová liniová vrstva s měsíčními údaji za 4 po sobě jdoucí roky. Na základě tvorby kategorizací vznikla třída prvků ve formátu ESRI geodatabase s pěti výslednými kategorizacemi. Všechny výsledné datové sady budou použitelné ve výuce i ke zpracování, pod podmínkou neporušení dohody o zapůjčení primárních dat a dohody o využití dat odvozených.

Časová analýza vybraných úseků

V rámci této analýzy museli být kvůli přílišnému množství úseků nejdříve definovány kategorie, které zúžili dostupné úseky na menší množství. Na základě toho vznikly analýzy dat formou dekompozice časových řad pro jednotlivé kategorie. Koridorové úseky, které obsahovali 9 úseků. Tyto úseky byly vybrány tak aby co nejlépe popisovali časový průběh počtu osobních vlakových souprav na čtyřech hlavních tranzitních koridorech v ČR. Hraniční přechody, které byly reprezentovány sedmi úseky v místech, kde hlavní tranzitní koridory protínají hranice ČR. Zde byl výběr motivován především analýzou průběhu počtu vlakových souprav, které proudí z a do ČR. Poslední kategorií byly zajímavé úseky, u kterých byl předpoklad výstředního chování z hlediska průběhu nákladní nebo osobní dopravy. Tyto kategorie byly popisovány na základě grafu měsíčních průběhů jednotlivých úseků, grafu trendu a grafu sezónní složky. Výsledkem této části jsou vypracované grafy dostupné v příloze (příloha 1-3) a interpretace, které jsou dostupné v kapitole hodnocení (viz kap. 6.4). Zejména v této části je pak možné také uvažovat jako výsledek samotný postup práce. Postup byl tvořen tak, aby bylo možné jej opakovat a v případě získání dat za další roky pokračovat v analýze.

Kategorizace stanic

V rámci této části byla realizována kategorizace na základě několika kombinací vstupních atributů a dvou metod. První metodou byl procentuální poměr typu dopravy v jednotlivých bodech zájmu rozdělený po 20 % s prostřední neutrální kategorií. Jako vstup při použití této metody, která dává relativní pohled na rozložení dopravy v ČR, byly použity roční úhrny počtu a hmotnosti vlakových souprav. Druhou metodou bylo shlukování. Po testování různých metod bylo nakonec, s přihlédnutím k rešerši odborných článků, vybráno hierarchické shlukování *Wardovou* metodou při použití *Euklidovské* metriky. Shlukovány byly nejdříve roční úhrny počty osobní a nákladních vlakových souprav a poté také hmotnost. Na závěr bylo přistoupeno ke shlukování měsíčních dat, při kterém byl kladen důraz na zahrnutí relativního průběhu roční křivky hmotnosti vlakových souprav. Výsledkem této části jsou mapové výstupy každého z pěti přístupů dostupné v příloze a interpretace dostupné v kapitole hodnocení (viz kap. 7.3).

9 DISKUZE

Zadání této diplomové práce vzniklo jako reakce na příslibená data o železniční infrastruktuře na podzim roku 2018. V tu chvíli nebylo příliš jasné, jak budou vypadat poskytnutá data, nebo zdali má SŽ k dispozici nějaká další data, která by šla použít v rámci analýzy. Vyhotoveno bylo proto zadání, které si dalo za cíl analyzovat data vytížení železniční infrastruktury pomocí kategorizace a analýzy časové složky za období let 2016 a 2017. Pokud jde o manipulaci s termínem vytížení železniční infrastruktury v rámci této práce, tak došlo k rozhodnutí, že vytížení je v této práci uvažováno jako zatížení železniční sítě zejména z hlediska počtu a hmotnosti vlakových souprav. Nelze jej chápat jako vytížení z hlediska limitní kapacity nebo počtu traťových kolejí. Bylo by nepochybně zajímavé použít data i tímto směrem, bohužel SŽ nemá momentálně k dispozici vhodnou datovou sadu, která by byla jednoduše dostupná k takovýmto typům analýz. Dále je také nutné zmínit, že v rámci rozsahu analýz bylo v zadání počítáno pouze s rozsahem dvou let, protože v tu chvíli to byl maximálně rozsah, který byl ze strany SŽ zaručený. Nakonec byly během řešení práce poskytnuty také datové sady, nejen za rok 2018, ale dokonce i za rok 2019. Vzhledem k tomu, že částí této diplomové práce je i analýza časových řad, tak bylo z hlediska aktuálnějších dat a delšího souvislého období žádoucí rozšířit řešený rozsah o tyto dva roky.

Při tvorbě zadání také nebylo příliš jasné, jakou strukturu budou nakonec poskytnutá data mít. Nakonec byly nejdříve poskytnuta data za úseky, která jsou tvořena počátečním a koncovým monitorovacím bodem za roky 2016 a 2017 (později také za roky 2018 a 2019). Původní plán bylo založit práci pouze na bodových datech za celé sledované období, která budou získána z dat úsekových. Tento plán bohužel nebyl možný, protože například ve stanicích, které mají funkci dopravního uzlu, tudíž se v nich protíná více směrů, nebylo vždy možné spočítat úhrn za celou stanici. Zjednodušeně prostý součet všech směrů nebyl možný, protože ne všechny úseky končily nebo naopak začínaly v uzlu a vlaky které uzlem projížděli, by byly započítány dvakrát. Podobný problém nastal, když stanice obsluhující pouze dva směry byla pro oba úseky pouze koncová, nebo pouze počáteční. Tudíž nebylo možné strojově spočítat hodnoty pro tuto stanici a jedinou volbou by bylo v tomto případě ručně počítat každou takovou stanici. Na základě vzniklých problémů byly nakonec poskytnuta i data za jednotlivé monitorovací body, bohužel pouze s výše zmiňovaným rozsahem 22 měsíců. Na základě těchto skutečností bylo nakonec přistoupeno k použití úsekových dat, jako vstupu pro časové analýzy, díky jejich časovému rozsahu. Data za jednotlivé monitorovací body byly na úkor mírné neaktuálnosti, kdy celý kompletní rok je pouze 2016, použity pro kategorizace, protože mají přímou vazbu na jednotlivé stanice a jiná dopravně a tarifně zajímavá místa. Na druhou stranu by bylo zajímavé z hlediska toho, že železniční síť je dělena na jednotlivé tratě, vytvořit kategorizace úsekových dat. Při stejné situaci, by tak byla zaručena aktuálnost této kategorizace. V případě této práce byla však zvolena cesta zpracování dat za jednotlivé monitorovací body.

Poskytnuté datové sady obsahovali v obou případech poměrně totožné atributy. Z hlediska číselných atributů byly použity pouze atributy v absolutních hodnotách, a to hmotnost a počet vlakových souprav. Byly vybrány proto, že z hlediska řešeného zatížení, se zdají být optimálními kandidáty a toho že ostatní relativní atributy nevnášejí do tohoto tématu nijak zásadní informace. Při použití hmotnosti a počtu vlakových souprav bez dělení dopravy sice vykazují poměrně velkou korelaci, ale při použití s rozdělením typu dopravy se situace mění a informační hodnota se zvyšuje. Nakonec

bylo této skutečnosti využito při kategorizacích. V kategorizacích nebyl použit ani dostupný kvalifikátor, který určoval názvem využití daného monitorovacího bodu, protože jednotlivých využití bylo celkově 37 a z velké části šlo o technické využití, protože jednotlivé monitorovací body nejsou ze své podstaty pouze železničními stanicemi, ale i dalšími tarifně a dopravně zajímavými místy.

V rámci kategorizace je mnoho možností a metod jaké lze zvolit. Z hlediska kartografického se nabízí například typizace. Kategorizace v této práci je zejména kvůli plánovanému využití metod dataminingu, založena na shlukování s doplňkem v podobě relativního rozdělení osobní a nákladní dopravy. Kartografická typizace by byla možným rozšířením této práce nad rámec jejího nynějšího rozsahu.

Při shlukování dat za jednotlivé monitorovací body bylo na základě rešerše zvoleno hierarchické shlukování. Při parametrizaci hierarchických metod shlukování je jedna z možností ukončení shlukovacího algoritmu číslo konečného počtu shluků. V rámci všech výsledných shlukování bylo toto číslo nastaveno na pět shluků, tedy jednotlivých kategorií, které byly následně interpretovány. Toto číslo bylo zvoleno jako optimální pro tuto práci s přihlédnutím na podrobnost a míru zobecnění. Pokud by byl požadavek na větší podrobnosti výsledných kategorií, nebo by byl problém s přílišným zobecněním, zejména početnějších shluků, bylo by možností konečný počet shluků navýšit. Další možností by též mohlo být odstranění odlehklých hodnot, které by byly definovány jako takové, které zasahují do jednoho nebo více shluků a zkreslují výsledky.

Během návrhu mapových výstupů bylo rozhodnuto, z hlediska toho, že nejde o hlavní výsledek této diplomové práce, pro použití mapového listu o velikosti A4. Symbologie jednotlivých kategorizací byla z tohoto důvodu nastavena jednoduší formou, než by dovolil větší formát. Aktuální situace zobrazuje jednotlivé kategorie tak, jak jsou pojmenovány v textu, rozdělené pouze barvou znaku. V tomto případě je celkový počet nebo hmotnost vlakových souprav pouze zobecněn intervalem v pojmenování kategorie. Při použití většího formátu mapového listu by bylo možné například toto zobecnění odstranit a použít pro počet a hmotnost více intervalů, které by byly zobrazeny velikostí mapového znaku, nebo použít více doplňujících textů, grafů nebo tabulek.

Největší příležitostí pro pokračování této práce je dle mého názoru shlukování relativních průběhů měsíčních úhrnů dat za jeden rok. Řešení testované v této práci (viz kap. 7.2.1), jako reakce na neuspokojivé výsledky shlukování absolutních měsíčních úhrnů hmotnosti vlakových souprav, podávalo v některých případech zajímavé výsledky. Zejména pak v případech shluku s velmi malým celkovým počtem monitorovacích bodů. Teoreticky by poté bylo možné přiřadit poměrně přesně stanice s podobným chováním v rámci roku. Problémy s velkým rozptylem v rámci jednoho shluku by bylo možné řešit například výběrem pouze takových shluků, které podávají dobré výsledky nebo filtrací vstupních dat tak, aby příliš odlehklé hodnoty nevstupovali do shlukování. Tento postup však nakonec nebyl v této práci použit, protože vytyčeným cílem byla kategorizace monitorovacích bodů za celé území ČR a kompletní dataset.

Dalším možným pokračováním této práce jsou predikce, zejména díky předpokladu, že data budou generována každý rok a tím bude vznikat stále delší časová řada. Výhodou je tak možnost nalézt takový algoritmus predikce, který by nejlépe fungoval na tato data s možností následné kontroly při vydání nových dat. Predikce je vrcholným výsledkem při komplexnějších analýzách časových řad. Výzvou, která je dána aktuální situací nouzového stavu v ČR z důvodu pandemie nemoci COVID-19, je analyzování těchto a podobných dat s časovou složkou. V rámci těchto dat by bylo možné na základě

měsíčních úhrnů poměrně přesně kvantifikovat dopady na železniční dopravu, a to jak osobní, tak nákladní.

Poskytnutá data jako taková byla anonymizována tak, aby mohla být zpracována a neobsahovala žádné citlivé údaje. O využití poskytnutých dat a dat, která jsou výsledkem této práce, bude jednáno s poskytovatelem (SŽ). Plánované je zejména využití ve výuce. Tato diplomová práce, která je postavena na získaných datech bude poskytnuta zástupcům SŽ jako výsledek spolupráce.

ZÁVĚR

Hlavním cílem této diplomové práce bylo provést analýzu časové složky dat využitosti železniční infrastruktury. Dále pak vyhodnocení dat o železniční infrastruktuře a jejich kategorizaci. Poskytnutá data za jednotlivé monitorovací body a úseky mezi nimi měla rozsah od roku 2016 do roku 2019. Data byla poskytnuta státní organizací Správa železnic.

Před samotnou praktickou částí této práce bylo nutné získat teoretický základ pro jednotlivé řešené oblasti. V první řadě šlo o legislativní rámec vývoje železnice v ČR a správce železniční dopravní cesty, které bylo nutné pochopit zejména z hlediska následné interpretace a pojmenování jednotlivých funkcí, které jsou vázány k problematice železnic v ČR. Další oblastí bylo položení teoretického základu k problematice železniční sítě v ČR, kde byl kladen důraz zejména na rozdělení železničních tratí a lokalizaci významných osobních a nákladních železničních koridorů. Teoretická část poté pokračovala průzkumem odborných publikací a článků zveřejněných na podobné téma, jako je řešené v této práci, nebo pracujících s daty o podobné povaze. Závěr teoretické části byl naplněn rešerší problematiky shlukování a časových řad, jakožto dvou metod, které jsou klíčové pro předkládanou diplomovou práci.

V rámci praktického řešení, ještě před samotnými analýzami, bylo nutné provést seznámení s daty a jejich předzpracování. Data nebyla v poskytnutém stavu vhodná pro okamžité použití a bylo nutné je upravit. Předzpracování dat proběhlo na základě požadavků daných analýzou časových řad a kategorizacemi. Data byla upravena do vhodné podoby filtrací, doplněním, odstraněním duplicit a kontingenčními tabulkami. Seznámení s daty proběhlo na základě popisu struktury a atributů dat pro jednotlivé monitorovací body a úseky mezi nimi. Dále byly vytvořeny jednoduché vizualizace formou základních mapových výstupů a grafů pro bližší pochopení situace zatížení železniční sítě v ČR.

První ze dvou provedených analýz nad poskytnutými daty byla analýza časových řad úsekových dat v časovém období 2016-2019. Analýza byla prováděna pomocí aditivního rozkladu na složky, pomocí kterých byly následně jednotlivé časové řady popisovány. Z důvodu velkého množství vstupních dat byly definovány 3 kategorie, jejichž zástupci byli poté podrobeni analýze. Byly to koridorové úseky, hraniční přechody a zajímavé úseky. Výsledek této analýzy jsou interpretace na základě vizuální analýzy grafů jednotlivých složek z rozkladu časových řad.

Druhou analýzou je kategorizace dat o železniční infrastruktuře za jednotlivé monitorovací body. Na počátku analýzy bylo definováno několik přístupů, na základě kterých byla následně data kategorizována. Přístupy byly definovány metodami kategorizace a kombinací vstupních atributů. První metodou bylo procentuální vyjádření poměru osobní a nákladní dopravy z hlediska počtu a hmotnosti vlakových souprav. Zde šlo o relativní vyjádření bez vlivu absolutních dat. Druhou metodou byla kategorizace formou shlukování. Na základě testování a rešerše bylo vybráno hierarchické shlukování *Wardovou* metodou s použitím *euklidovské* metriky. Vstupními atributy byly roční úhrny dat o počtu nebo hmotnosti osobních a nákladních vlakových souprav a měsíční úhrny hmotnosti vlakových souprav celkem. Použití těchto dat nevykazovalo požadované výsledky, a tak bylo přistoupeno k testování dalších variant, tak, aby výsledky co nejvíce refletovali relativní kolísání časových řad. Výsledkem této

části je datová sada kategorizovaných dat, mapové výstupy všech kategorizací a interpretace s hodnocením jednotlivých shluků.

Tato práce je pilotním příkladem zpracování těchto unikátních dat o vytížení železniční infrastruktury v takovémto formátu, která byla takto uvolněna zcela poprvé. Výsledné analýzy nabízejí zajímavý pohled na situaci železniční dopravy na sítích SŽ v ČR. Kompletní pohled na variabilitu monitorovacích bodů dávají kategorizace, u kterých je v rámci práce nabídnuto hned několik variant, které jsou náležitě okomentovány. Bližší pohled na časovou variabilitu vybraných kategorií stanic ze čtyř po sobě jdoucích let, dává analýza časových řad. Zde je situace chování v rámci tohoto časového období popsána s přesností na jednotlivé měsíce v rámci čtyř let, s důrazem kladeným zejména na hlavní tranzitní koridory, hraniční přechody a dopravně zajímavé monitorovací body. Datové sady, vzniklé jako výsledek předkládané diplomové práce, budou dostupné pro další zpracování na katedře geoinformatiky, tak aby nebyla porušena dohoda s poskytovatelem původních dat. Výsledky práce budou představeny SŽ.

POUŽITÁ LITERATURA A INFORMAČNÍ ZDROJE

BRINKE, Josef, 1999. *Úvod do geografie dopravy*. Praha: Karolinum. ISBN 80-718-4923-5.

BUDILOVÁ, Lenka, 2019. *Analýza nákladní železniční dopravy v oblasti Jesenicka* [online]. [cit. 2019-10-11]. Dostupné z: https://dspace.cvut.cz/bitstream/handle/10467/85353/F6-BP-2019-Budilova-Lenka-Analyza_nakladni_zeleznicni_doprava_Jesenicko.pdf?sequence=-1&isAllowed=y. Bakalářská práce. ČVUT. Vedoucí práce Ing. Zdeněk Michl.

ČSÚ, 2011. *Rozloha území a počet obyvatel České republiky podle krajů a okresů k 1. 1. 2011* [online]. [cit. 2019-02-23]. Dostupné z: <https://www.czso.cz/documents/10180/20548145/4032120101.pdf/c1675e94-73a8-490f-ac08-08a45abb6c9b?version=1.0>

DOBEŠOVÁ, Zdena., KUČERA, Michal. *Time series of workload on railway routes*, In Silhavy, R. (ed.) CSOC 2019. Artificial Intelligence and Algorithms in Intelligent Systems, book series Advances in Intelligent Systems and Computing, vol. 985, pp. 370-380, 2020, Springer, Cham. DOI: 10.1007/978-3-030-19810-7_37

GAŠPARÍK, Jozef a Jiří KOLÁŘ, 2017. *Železniční doprava: technologie, řízení, grafikony a dalších 100 zajímavostí*. Praha: Grada Publishing. ISBN 978-80-271-0058-3.

HANČLOVÁ, Jana a Lubor TVRDÝ, 2003. *Úvod do analýzy časových řad*. Ostrava. Ekonomická fakulta, VŠB-TU.

Historie železnice v ČR: Historie našich železnic v kostce, 2019. SŽDC [online]. 2019 [cit. 2019-05-12]. Dostupné z: <https://www.szdc.cz/documents/50004227/50156852/historie-zeleznice-v-cr.pdf/168d4de0-53ab-4b38-8ecb-db84f84dd68f?version=1.0>

Informace o kategorizaci železniční sítě, 2019. Ministerstvo dopravy ČR [online]. [cit. 2019-02-20]. Dostupné z: <https://www.mdcrcz.cz/Dokumenty/Drazni-doprava/Zeleznicni-infrastruktura/Informace-o-kategorizaci-zeleznicni-site>

KLÍMEK, Petr, 2008. Shlukovací metody v Data Miningu. *E + M Ekonomie a Management* [online]. Liberec: Technická univerzita v Liberci, 2008, 2008(2), 120–126 [cit. 2019-04-11]. Dostupné z: http://www.ekonomie-management.cz/download/1331826675_2e7a/11_klimek.pdf

KLÍMEK, Petr a Roman KASAL, 2007. *Počítačové zpracování dat v programu Statistica: studijní pomůcka pro distanční studium*. Zlín: Univerzita Tomáše Bati ve Zlíně. ISBN 978-80-7318-526-8.

Komentář ke studii „Kritéria kategorizace železniční sítě“, 2019. *Ministerstvo dopravy ČR* [online]. [cit. 2019-02-20]. Dostupné z: https://www.mdcr.cz/getattachment/Dokumenty/Drazni-doprava/Zeleznicni-infrastruktura/Informace-o-kategorizaci-zeleznicni-site/Komentar-ke-studii-Kriteria-kategorizace-zeleznicni-site_.doc.aspx

KŘIVÝ, Ivan, 2012. *Analýza časových řad* [online]. Ostrava: Ostravská univerzita v Ostravě [cit. 2019-03-21]. Dostupné z: <http://www1.osu.cz/~bujok/files/ancas.pdf>

KUČERA, Jiří. *Metody kategorizace dat*. 2008. Praha.

KUČERA, Jiří, 2008. *Metody kategorizace dat*. Brno. Dostupné také z: https://is.muni.cz/th/w8lgz/Metody_kategorizace_dat.pdf. Bakalářská práce. Masarykova univerzita. Vedoucí práce Mgr. Matěj Štefaník.

LITSCHMANNOVÁ, Martina, 2010. *Úvod do analýzy časových řad*. Ostrava: VŠB – TU Ostrava.

Microsoft, 2019. *Vytvoření kontingenční tabulky k analýze dat listu* [online]. [cit. 2019-02-10]. Dostupné z: <https://support.office.com/cs-cz/article/vytvo%C5%99en%C3%AD-kontingen%C4%8Dn%C3%AD-tabulky-k-anal%C3%BDze-dat-listu-a9a84538-bfe9-40a9-a8e9-f99134456576>

MIRVALD, Stanislav, 2000. *Geografie dopravy II: silniční a železniční doprava*. Plzeň: Západočeská univerzita. ISBN 80-708-2673-8.

Nariadení Evropského parlamentu a Rady (EU) č. 913/2012 o evropské železniční síti pro konkurenceschopnou nákladní dopravu. In: *Úřední věstník Evropské unie*. L 276, 20. 10. 2010, s. 22-32. Dostupné také z: <https://eur-lex.europa.eu/legal-content/CS/TXT/PDF/?uri=CELEX:32010R0913&from=EN>

O nás: Vznik SŽDC, 2020. *SŽDC* [online]. 2020 [cit. 2020-05-12]. Dostupné z: <https://www.szdc.cz/o-nas/vse-o-sprave-zeleznic/vznik-szdc>

ROUBALÍK, Štěpán, 2017. *Analýza toků veřejné dopravy na základě dat z mýtných bran* [online]. Olomouc [cit. 2019-03-11]. Dostupné z: <http://www.geoinformatics.upol.cz/dprace/bakalarske/roubalik17/documents/roubalikBP.pdf>. Bakalářská práce. Univerzita Palackého v Olomouci. Vedoucí práce RNDr. Jaroslav BURIAN, Ph.D.

Silniční doprava: pozemní komunikace, veřejné služby v přepravě cestujících, dráhy: redakční uzávěrka k.., 2011-. Ostrava: Sagit. ÚZ.

Statistická ročenka 2019, 2020. In: *Správa železnic* [online]. [cit. 2020-03-15].
Dostupné z:

<https://www.szdc.cz/documents/50004227/64057801/Statistick%C3%A1+ro%C4%8Denka+2019/2212bd66-c7e7-44d6-9a5b-f337df90f5eb>

ŠARMANOVÁ, Jana, 2012. *Metody analýzy dat*. Ostrava. ISBN 978-80-248-2565-6.

ŠIROKÝ, Jaromír, 2018. *Technologie dopravy*. Čtvrté doplněné vydání.
Pardubice: Univerzita Pardubice. ISBN 978-80-7560-159-9.

ŠLACHTOVÁ, Hana, Jiří MICHALÍK a Ondřej VOLF, 2007. Zpráva o zpracování Strategické hlukové mapy ČR. In: *MZČR* [online]. Ostrava, 2007 [cit. 2019-03-11]. Dostupné z:

http://www.mzcr.cz/hlukovemapy/Soubor.ashx?souborID=24694&typ=application/pdf&nazev=Zprava_o_zpracovani_Strategicke_hlukove_mapy_CR.pdf

Tiskové prohlášení SŽ: Ze Správy železniční dopravní cesty je od 1. ledna 2020 Správa železnic, 2020. *SŽDC* [online]. Praha, 2020 [cit. 2020-05-12]. Dostupné z:

<https://www.szdc.cz/documents/50004227/84337901/tiskov%C3%A1+zpr%C3%A1va+ke+sta%C5%BEen%C3%AD/fc9115b6-be10-437e-a53f-dbfdf78aa889>

Tranzitní železniční koridory, 2019. *Ministerstvo dopravy ČR* [online]. [cit. 2019-02-20]. Dostupné z: <https://www.mdcr.cz/Dokumenty/Drazni-doprava/Zeleznicni-infrastruktura/Tranzitni-zeleznicni-koridory>

Zákon č. 266/1994 Sb., o dráhách, 1994. In: *Sbírka zákonů České republiky*. částka 79. Dostupné také z: <https://www.zakonyprolidi.cz/cs/1994-266>

PŘÍLOHY

SEZNAM PŘÍLOH

Vázané přílohy:

- Příloha 1 Grafy rozkladu časových řad pro vybrané úseky na tranzitních koridorech
- Příloha 4 Mapa kategorizace na základě podílu osobní a nákladní dopravy A1
- Příloha 5 Mapa kategorizace na základě podílu osobní a nákladní dopravy A2
- Příloha 6 Mapa kategorizace na základě shlukování ročních dat B1
- Příloha 7 Mapa kategorizace na základě shlukování ročních dat B2
- Příloha 8 Mapa kategorizace na základě shlukování ročních dat C1

Elektronické přílohy

- Příloha 2 Grafy rozkladu časových řad pro vybrané hraniční úseky
- Příloha 3 Grafy rozkladu časových řad pro vybrané zajímavé úseky

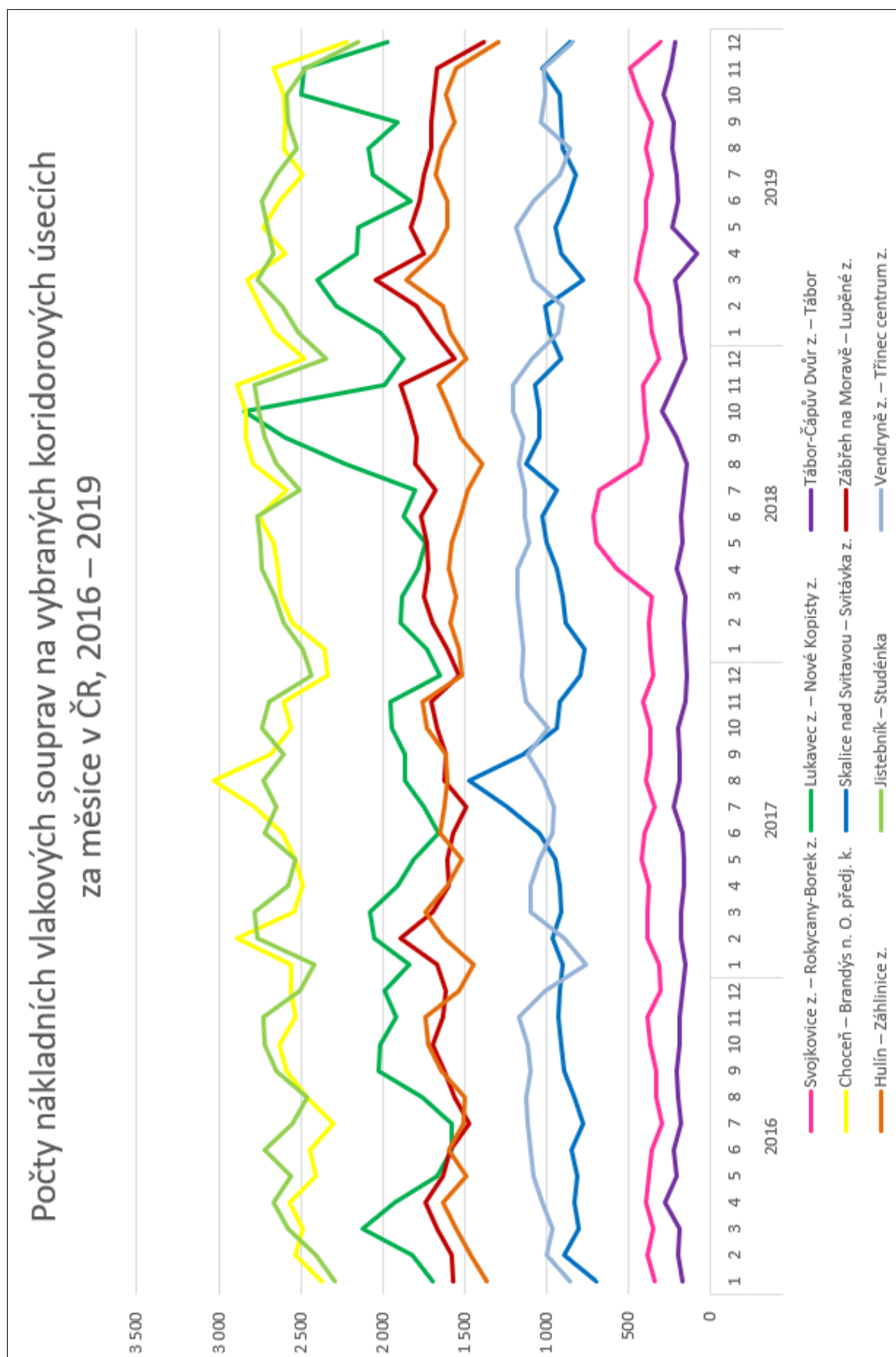
Volné přílohy

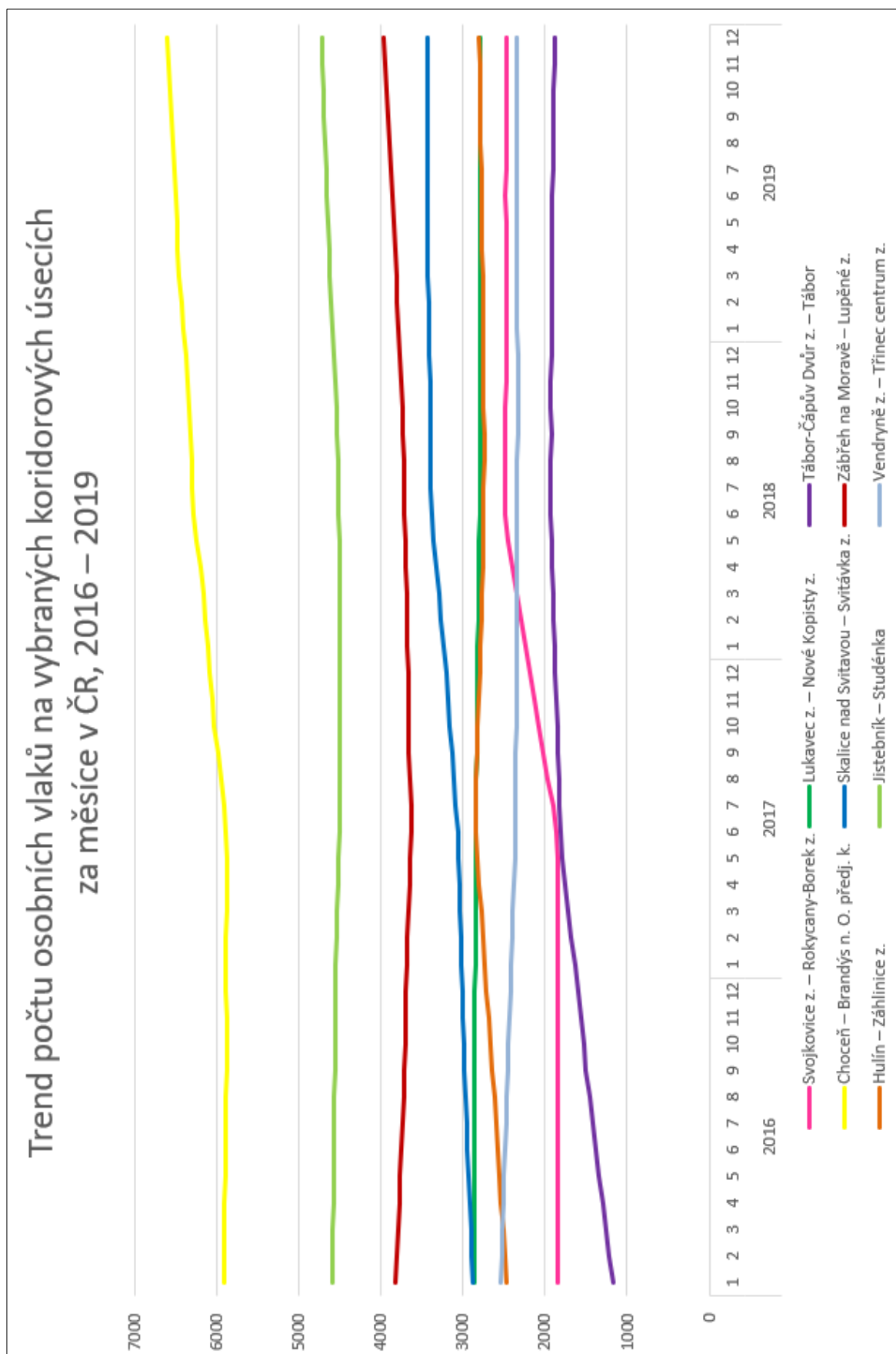
- Příloha 9 Poster
- Příloha 10 DVD

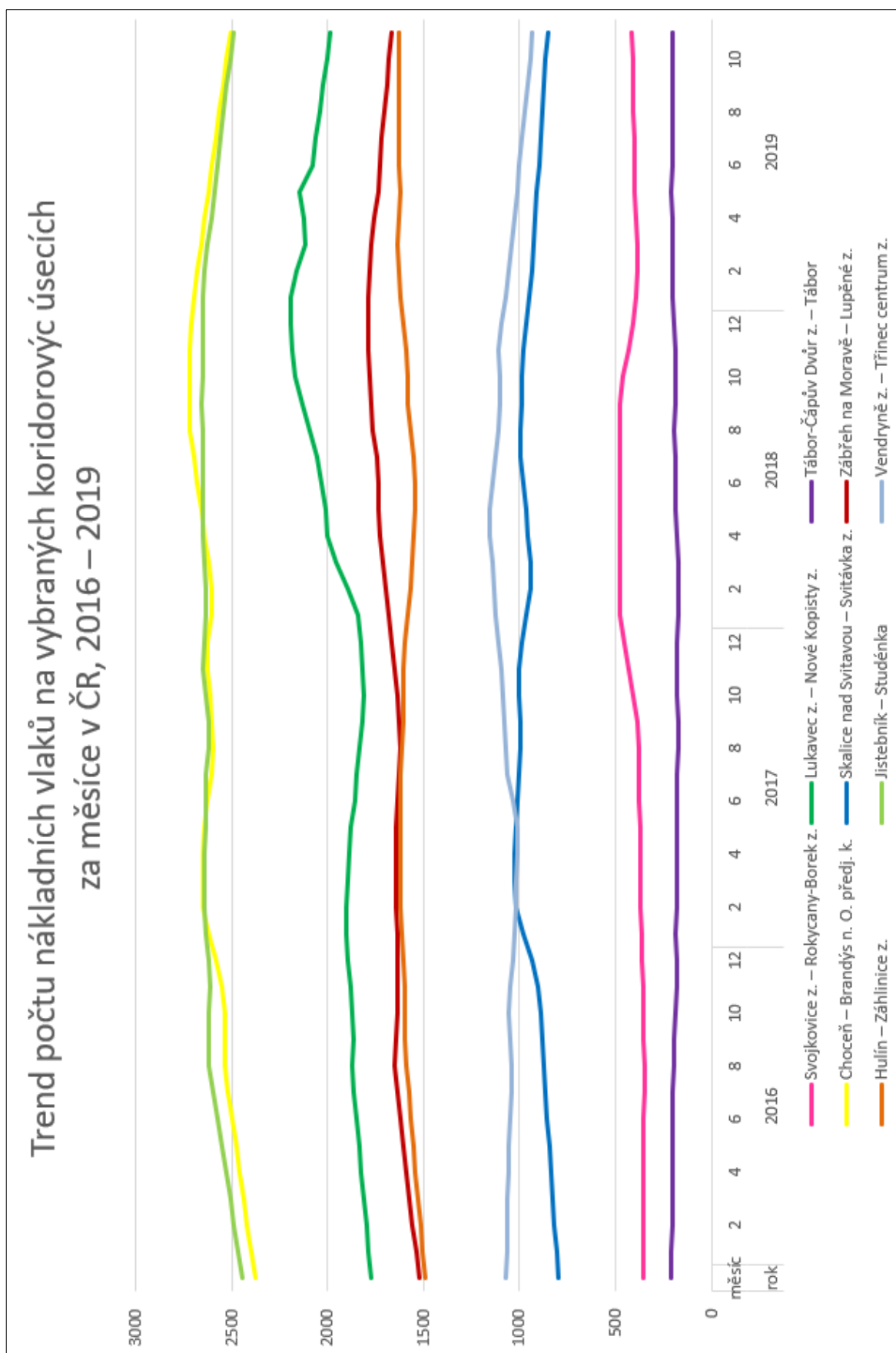
Popis struktury DVD

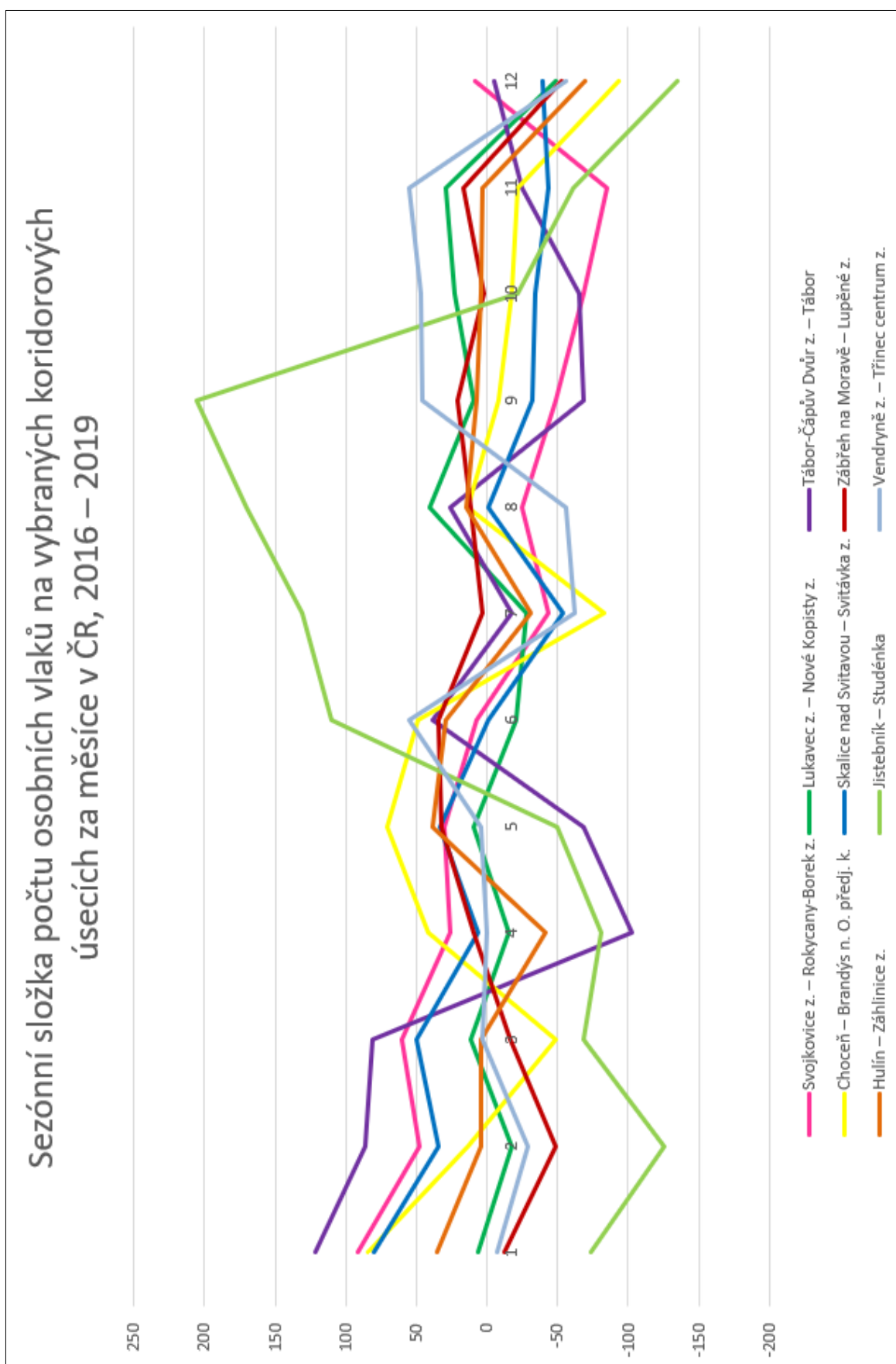
Adresáře:

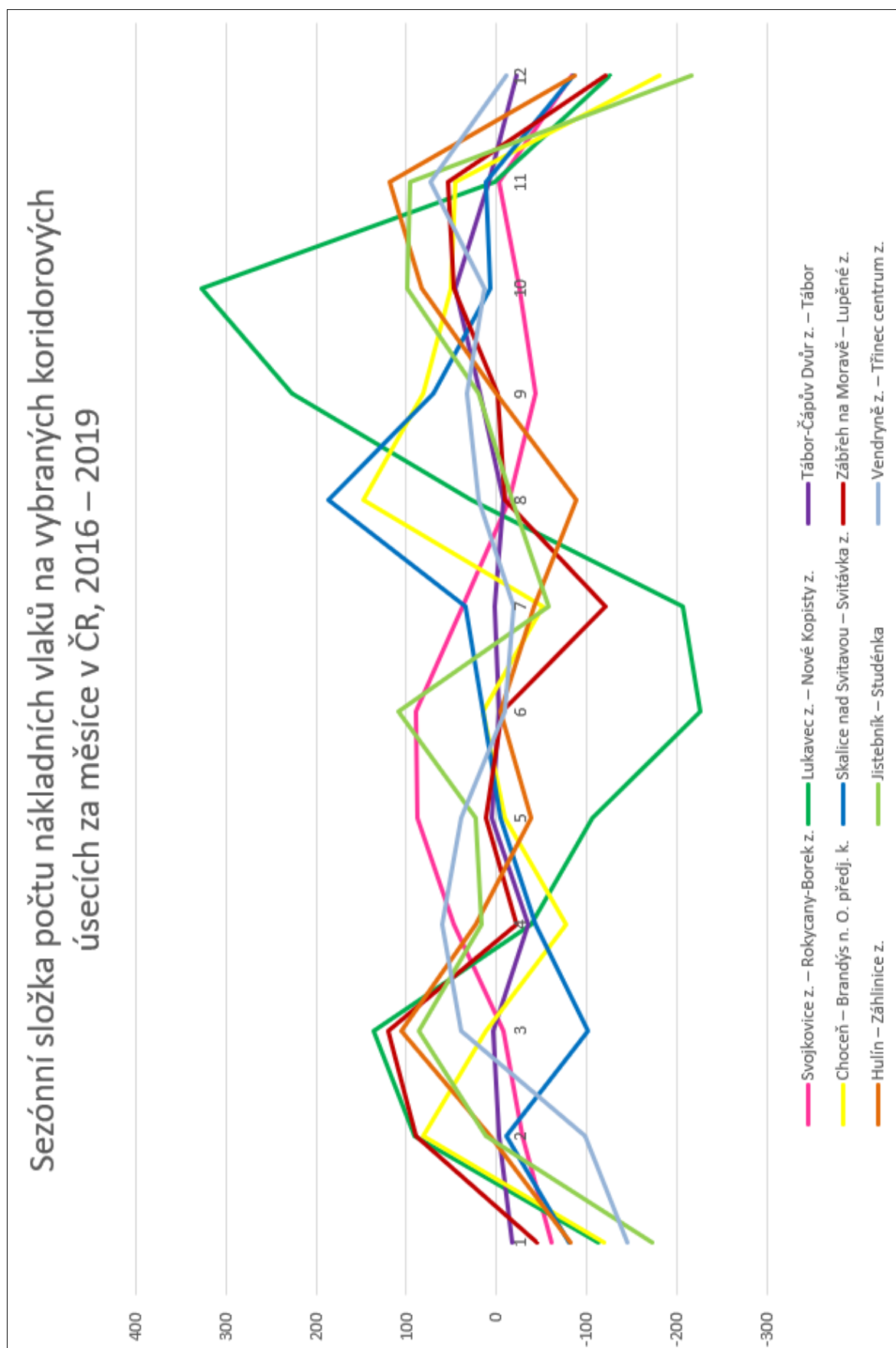
- 1_text_prace
- 2_web
- 3_vystupni_data
- 4_prilohy









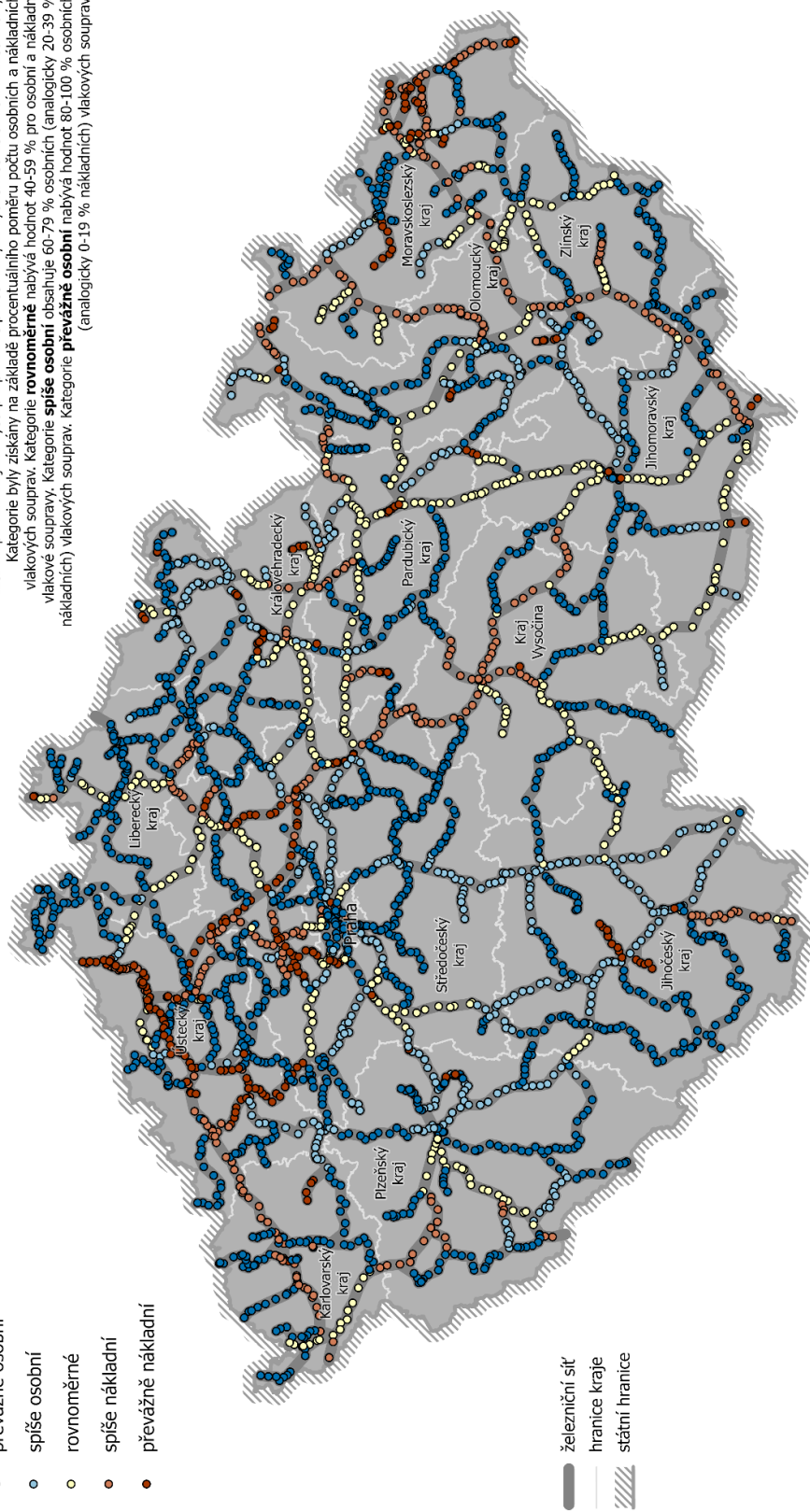


KATEGORIZACE ŽELEZNIČNÍCH STANIC, DOPRAVNĚ ZAJÍMAVÝCH A TARIFNÍCH MÍST Z ČÍSELNÍKU SR70 V ČR V ROCE 2016

A2. Procentuální vyjádření hmotnosti osobních a nákladních vlakových souprav za rok

Tato mapa vznikla jako výstup diplomové práce Analýza vyžitosti železniční infrastruktury. Kategorie byly získány na základě procentuálního poměru počtu osobních a nákladních vlakových souprav. Kategorie **rovnoměrné** nabývá hodnot 40-59 % pro osobní a nákladní vlakové soupravy. Kategorie **spíše osobní** obsahuje 60-79 % osobních (analogicky 20-39 % nákladních) vlakových souprav. Kategorie **převážně osobní** nabývá hodnot 80-100 % osobních (analogicky 0-19 % nákladních) vlakových souprav.

- převážně osobní
- spíše osobní
- rovnoměrné
- spíše nákladní
- převážně nákladní

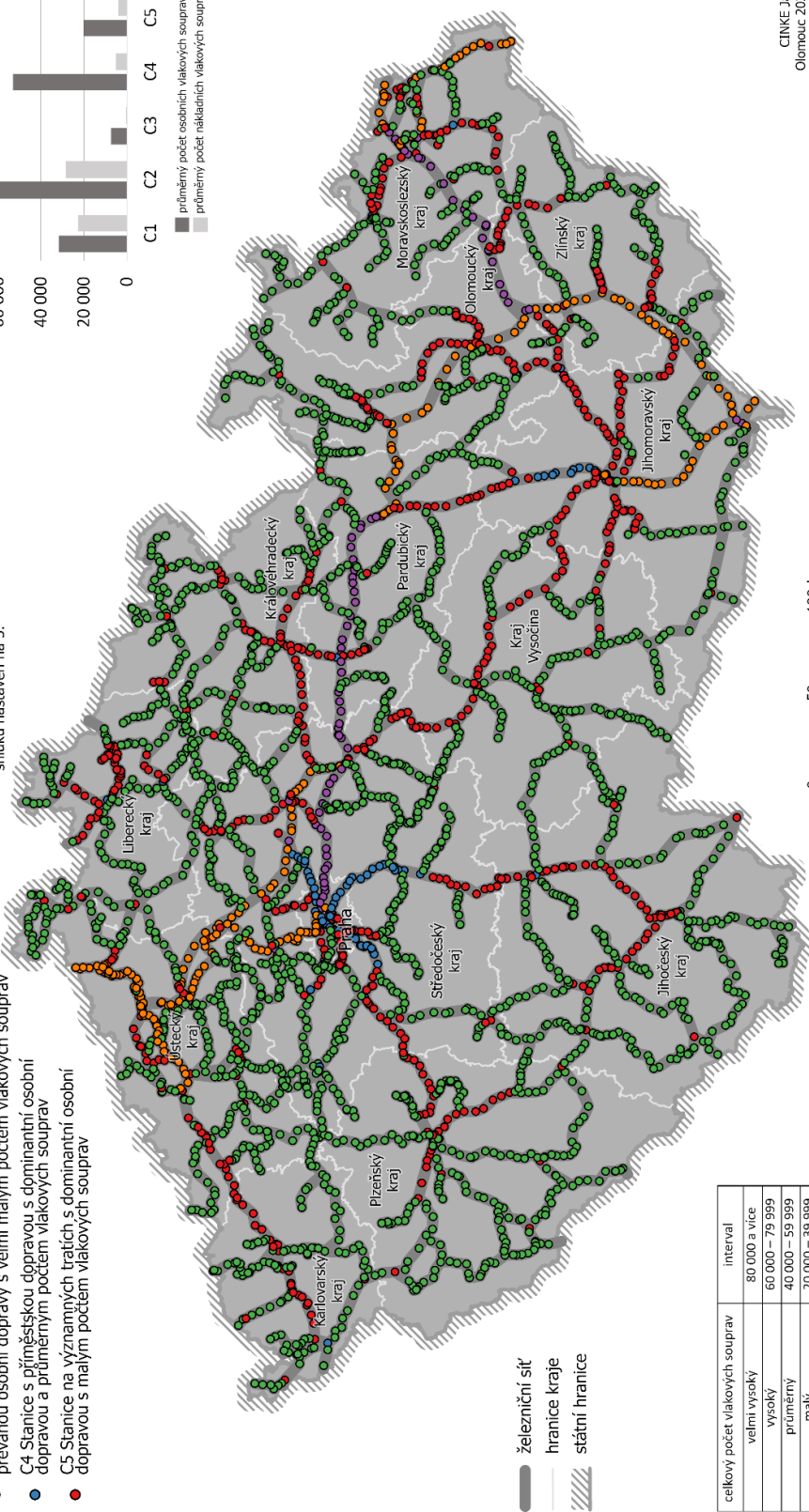
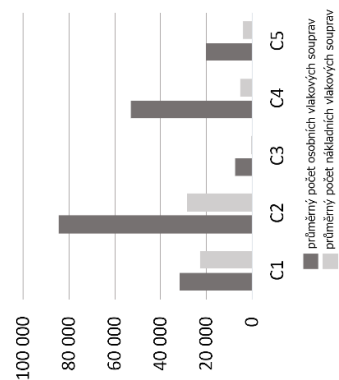


KATEGORIZACE ŽELEZNIČNÍCH STANIC, DOPRAVNĚ ZAJÍMAVÝCH A TARIFŇNÍCH MÍST Z ČÍSELNÍKU SR70 V ČR V ROCE 2016

B1. Počet osobních a nákladních vlakových souprav za rok

- C1 Stanice na nákladních a osobních koridorech s mírnou převahou osobní dopravy a průměrným počtem vlakových souprav
- C2 Stanice na tranzitních koridorech s vysokou převahou osobní dopravy a velmi vysokým počtem vlakových souprav
- C3 Stanice na regionálních a nadregionálních tratích s dominantní převahou osobní dopravy s velmi malým počtem vlakových souprav
- C4 Stanice s příměstskou dopravou s dominantní osobní dopravou a průměrným počtem vlakových souprav
- C5 Stanice na významných tratích s dominantní osobní dopravou s malým počtem vlakových souprav

Tato mapa vznikla jako výstup diplomové práce. Analýza vyřazenosti železniční infrastruktury, kategorie byly získány na základě shlukování atributů počtu osobních a nákladních vlakových souprav. Bylo použito hierarchické shlukování, Euklidovská metrika a Wardova metoda, kde byl počet shluků nastaven na 5.



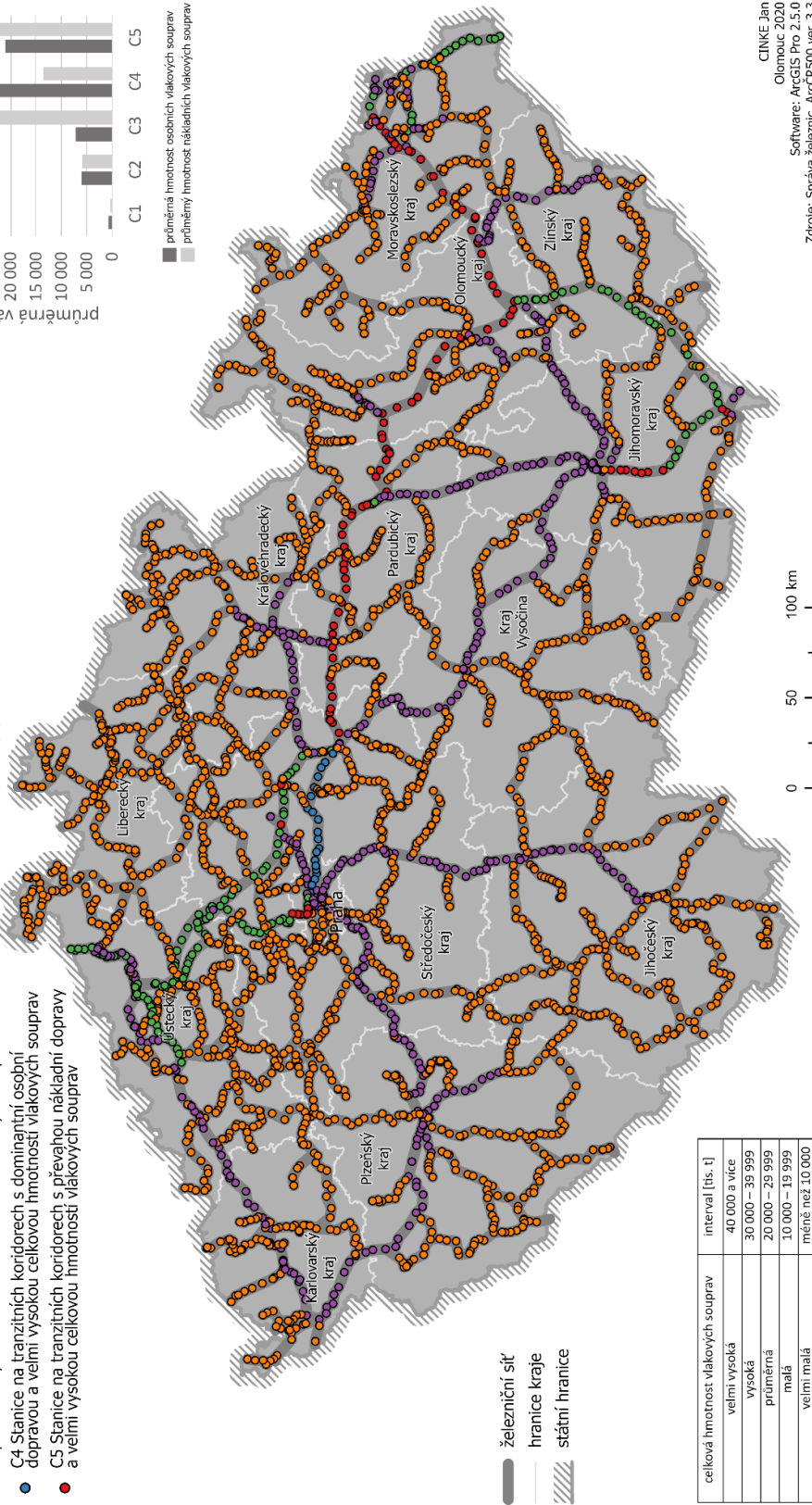
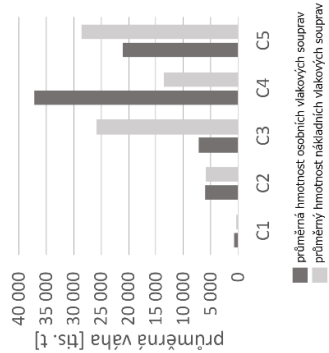
celkový počet vlakových souprav	interval
velmi vysoký	80 000 a více
vysoký	60 000 – 79 999
průměrný	40 000 – 59 999
malý	20 000 – 39 999
velmi malý	méně než 20 000

KATEGORIZACE ŽELEZNIČNÍCH STANIC, DOPRAVNĚ ZAJÍMAVÝCH A TARIFNÍCH MÍST Z ČÍSELNÍKU SR70 V ČR V ROCE 2016

B2. Hmotnost osobních a nákladních vlakových souprav za rok

- C1 Stanice na regionálních a nadregionálních tratích s převahou osobní dopravy a velmi malou celkovou hmotností vlakových souprav
- C2 Stanice na osobních a nákladních koridorech a významných tratích s nepatrnou převahou osobní dopravy a malou celkovou hmotností vlakových souprav
- C3 Stanice na osobních a nákladních koridorech s dominantní nákladní dopravou a vysokou celkovou hmotností vlakových souprav
- C4 Stanice na tranzitních koridorech s dominantní osobní dopravou a velmi vysokou celkovou hmotností vlakových souprav
- C5 Stanice na tranzitních koridorech s převahou nákladní dopravy a velmi vysokou celkovou hmotností vlakových souprav

Tato mapa vznikla jako výstup diplomové práce Analýza vyřazenosti železniční infrastruktury. Kategorie byly získány na základě shlukování atributů hmotnosti osobních a nákladních vlakových souprav. Bylo použito hierarchické shlukování, Euklidovská metrika a Wardova metoda, kde byl počet shluků nastaven na 5.



celková hmotnost vlakových souprav	interval [tis. t]
velmi vysoká	40 000 a více
vysoká	30 000 – 39 999
průměrná	20 000 – 29 999
malá	10 000 – 19 999
velmi malá	méně než 10 000

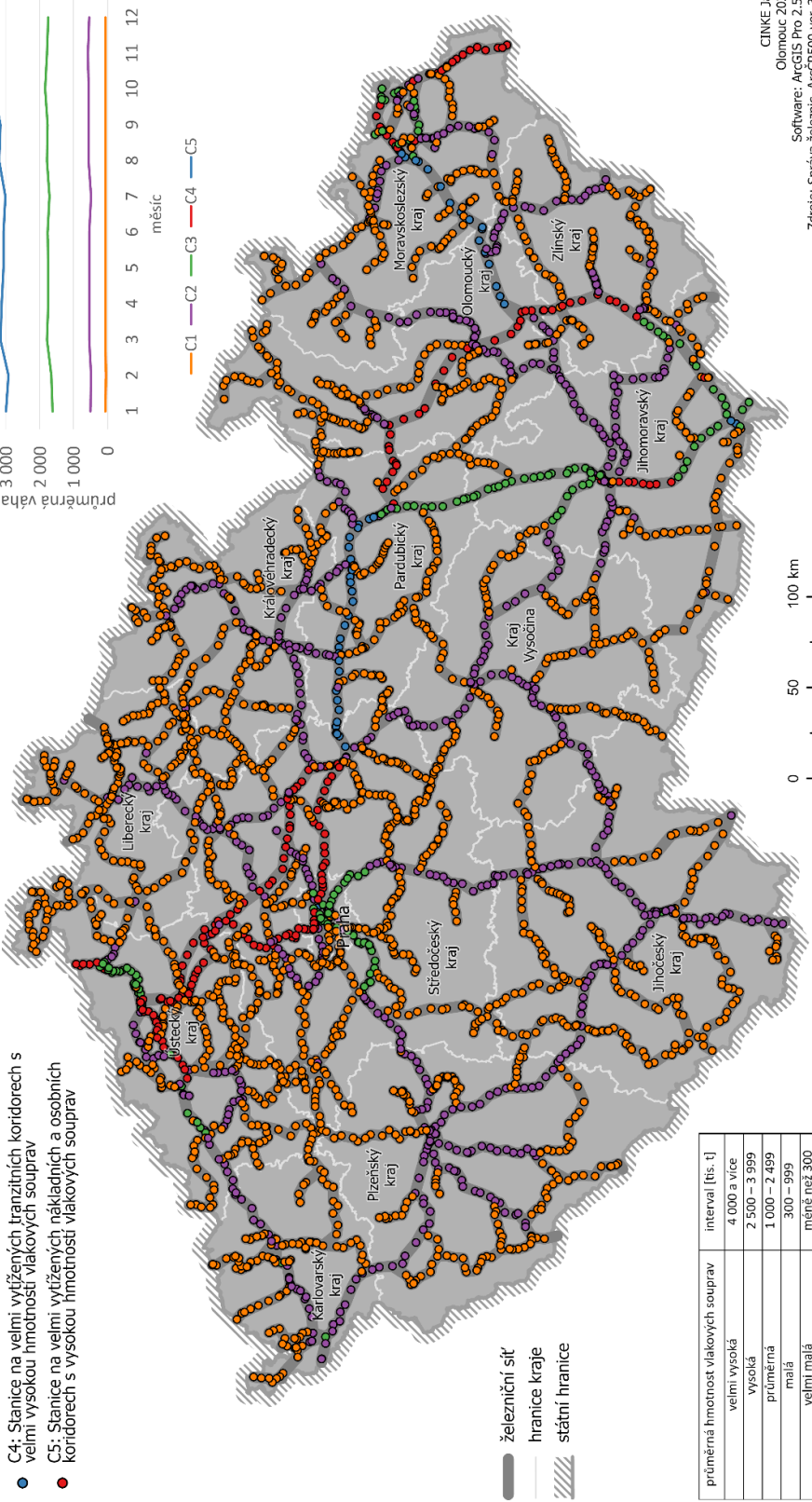
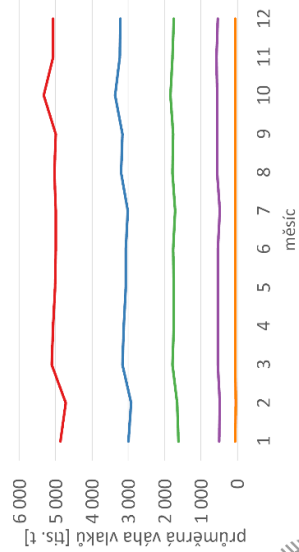
- železniční síť
- hranice kraje
- státní hranice

KATEGORIZACE ŽELEZNIČNÍCH STANIC, DOPRAVNĚ ZAJÍMAVÝCH A TARIFNIČNÍCH MÍST Z ČÍSELNÍKU SR70 V ČR V ROCE 2016

C1. Hmotnost osobních a nákladních vlakových souprav za měsíce

- C1: Stanice na regionálních a méně vytižených nadregionálních tratích s velmi malou hmotností vlakových souprav
- C2: Stanice na méně vytižených koridorech a vytižených nadregionálních tratích s malou hmotností vlakových souprav
- C3: Stanice na vytižených tranzitních koridorech a nadregionálních tratích s průměrnou hmotností vlakových souprav
- C4: Stanice na velmi vytižených tranzitních koridorech s velmi vysokou hmotností vlakových souprav
- C5: Stanice na velmi vytižených nákladních a osobních koridorech s vysokou hmotností vlakových souprav

Tato mapa vznikla jako výstup diplomové práce Analýza vytiženosti železniční infrastruktury. Kategorie byly získány na základě shlukování atributů hmotnosti osobních a nákladních vlakových souprav celkem, ve 12 měsíčních hodnotách. Bylo použito hierarchické shlukování, Euklidovská metrika a Wardova metoda, kde byl počet shluků nastaven na pět.



průměrná hmotnost vlakových souprav	interval [tis. t]
velmi vysoká	4 000 a více
vysoká	2 500 – 3 999
průměrná	1 000 – 2 499
malá	300 – 999
velmi malá	méně než 300