

Prostorové chování studentů Institutu geoinformatiky

Petra Linhartová

Institut geoinformatiky, Hornicko-geologická fakulta,
Vysoká škola báňská – Technická univerzita Ostrava,
17. listopadu 15/2172, 708 33, Ostrava-Poruba, Česká republika
petra.linhartova@vsb.cz

Abstrakt. Bakalářská práce se zabývá možnostmi analýzy pohybu vybrané skupiny respondentů pomocí mobilního zařízení, které přes aplikaci Časová osa kontinuálně zaznamenává geografickou polohu. Dále řeší otázku citlivých a osobních údajů respondentů v souladu se Zákonem o ochraně osobních údajů. Časové a prostorové analýzy jsou prováděny odděleně a výsledkem je vzhledem k slabé účasti spíše návod či ukázání možností zpracování časoprostorových dat než skutečné statistické vyhodnocení prostorového pohybu studentů.

Klíčová slova: prostorové chování, časoprostorová analýza, ochrana osobních údajů.

Abstract. Spatial Behaviour of Students' Journeys – Example of the Institute of Geoinformatics. This bachelor thesis deals with possibilities of analyzing space-time moves of a representative group by mobile device which continuously records geographic location by Timeline application. It discusses the topic of personal and delicate data in compliance with Act No. 101/2000 Coll., on the Protection of Personal Data. Spatio-temporal analyses are executed separately and the result is, considering the weak turnout, more of a manual or a demonstration of possibilities of processing space-time data than the real statistical evaluation of spatial behavior of students.

Keywords: Spatial Behaviour, Spatio-temporal Analysis, the Protection of Personal Data

1 Úvod

Prostorové chování se sleduje již dlouhou dobu, sledují se pohyby zvířat, migrace lidí a mobilita lidí v dopravě apod. Společně s rozvojem mobilních technologií a dostupností geodat a časoprostorových dat z mobilních zařízení jde i rozvoj analytických nástrojů pro jejich zpracování a interpretaci.

2 Použitá data a software

Data. Hlavní složkou dat jsou data nasbíraná respondenty, dále pak bodová vrstva Points of Interests se zájmovými body upravená podle předpokladu navštívení studentem VŠ. Nedílnou součástí je polygonová vrstva obcí ČR použitá pro lepší orientaci a vrstva adresních bodů města Ostravy.

Software. Pro základní zpracování dat byly použity programy ArcMap 10.3, Quantum GIS 2.10.1, MS Excel. Prostorové analýzy vznikaly v opensource software GeoDa. Časoprostorové analýzy a animace

3 Sběr dat

Pomocí mobilních telefonů se sbírají různá časoprostorová data týkající se pohybu osob pro komerční i vědecké účely. [1] Sběr se dá provádět dvěma způsoby a to pasivní nebo aktivní metodou. Pasivní metodou rozumíme již existující lokalizační údaje vzniklé pomocí aktivity na mobilním telefonu (SMS, hovory a další služby), nevýhodou je nepravidelnost dat v čase a složitější zpracování. Aktivní metoda je ta, kdy se telefon pravidelně dotazuje na svoji polohu a zaznamenává se tak poměrně přesná trajektorie pohybu uživatele zařízení.[1] Data jsou dnes mnohem přístupnější a v návaznosti na jejich dostupnost je také rozvoj analytických nástrojů pro zpracování a rozbor takových dat. Díky rozšíření mobilních telefonů mezi populací je možno sledovat tzv. prostorovou mobilitu obyvatel a tím získávat údaje pro provozovatele veřejné dopravy, marketingové společnosti, sociology či státní útvary.

Pro účely mé bakalářské práce bylo potřeba získat prostorová data studentů geoinformatiky. Oslovila jsem tedy své spolužáky od prvních po páté ročníky emailem, ve kterém jsem je žádala o zapojení do sběru dat. Vysvětlila jsem, jakým způsobem se budou data sbírat, a co všechno je k tomu třeba udělat.

Pro sběr dat jsem využila službu společnosti Google Inc. Časová osa (Timeline). Tato služba zaznamenává polohu uživatele mobilního telefonu s operačním systémem Android 4.04 a vyšší popř. iOS. Patří tedy mezi aktivní lokalizační metody. Dále ukazuje podle naměřených bodů přibližnou trasu pohybu a přichytávání k zájmovým bodům. Přesnost vykreslení trasy závisí na rychlosti pohybu mobilního telefonu, možnostech a kvalitě připojení k internetu (wi-fi, mobilní internet) a dalších faktorech. V ideálním případě je poloha zaznamenávána pravidelně, co dvě minuty.

Z časové osy je možné data stáhnout ve formátu KML (Keyhole Markup Language). Při zobrazení dat v tomto formátu vidíme pouze jedinou polylinii, která neobsahuje časové značky zaměření lomových ani koncových bodů.

4 Anonymizace dat

Pro zpracování dat spojených s polohou fyzických osob je nutné tato data anonymizovat, protože zobrazují osobní nebo citlivé údaje jako je například bydliště respondent

Zákon o ochraně osobních údajů. Citlivé údaje týkající se fyzických osob mohou být spravovány pouze tak, aby odpovídaly Zákonu č. 101/2000 Sb., o ochraně osobních údajů v jeho účinném znění. Při vytváření své bakalářské práce pracuji s daty týkajícími se polohy osob, ze kterých lze jednoznačně určit studentovo bydliště a další možné citlivé údaje. Proto je nezbytná anonymizace dat pro možnost zpracování a následného zveřejnění výsledků mé bakalářské práce.

Zpracovávaná data tedy mohou obsahovat jak osobní, tak citlivé údaje a proto s jejich zpracováním fyzickou osobou musí subjekt, tedy osoba, které se tyto údaje týkají, souhlasit. Následná úprava dat na formu takovou, kde nelze o údajích mluvit jako o údajích osobních či citlivých, je prostředkem pro získání dat anonymních, čili těch, které nelze spojit s konkrétní osobou.

4.1 Kroky anonymizace

Anonymní předání dat. Sběr dat od respondentů proběhl ihned po skončení měření 14. 10. 2015. Data byla respondenty uložena na internetové úložiště Google Disk, které bylo přístupné pouze respondentům a mně. Z dat tak není patrné, kým byla na úložiště vložena.

Agregace bodů na čtvercovou síť. Hlavním krokem anonymizace dat je převedení naměřených bodů na čtvercovou síť kvůli možnosti dedukce citlivých údajů z dat od jednotlivých respondentů. K zamezení možnosti rozeznání místa bydliště nebo získání podobné informace bylo nutné zjistit vhodnou velikost buňky čtvercové sítě. Tedy provést analýzu hustoty adresních bodů v městě Ostravě. Agregaci použiji až pro zobrazení samotných výsledků, protože pro některé prostorové a časové analýzy je vhodnější bodová či liniová vrstva.

Statistika EDA nad adresními body města Ostravy. Pro zjištění, jak velká má být buňka čtvercové sítě velká, jsem provedla statistiku bodů nad adresními body města Ostravy. Vytvořila jsem si vrstvy čtvercové sítě pomocí funkce Fishnet v programu ArcMap 10.3 s buňkou o hraně v různých velikostech a to 100, 200, 250 a 500 m. Každou čtvercovou síť jsem pomocí funkce Spatial Join spojila s adresními body města Ostravy. Následně jsem v atributové tabulce pomocí nástroje Summarize vytvořila výstupní tabulku, kde v jednom sloupci bylo FID buňky sítě a v druhém počet bodů spadajících do buňky. Nad touto tabulkou jsem provedla v MS Excel statistiku EDA pro zjištění různých středních hodnot.

Tabulka 1. Statistika EDA nad adresními body

Velikost strany v m	Průměr	Medián	Směrodatná odchylna
500	46,56	29,00	48,76
250	16,45	11,00	15,86
200	11,81	8,00	11,05
100	4,42	3,00	3,50

Jako nejvhodnější buňku čtvercové sítě jsem zvolila tu se stranou o délce 250 metrů. Z průměrného počtu 16 adresních bodů se již rozhodně nedá jednoznačně určit, na kterém adresním bodu by se mohlo nacházet bydliště respondenta.

5 Zpracování dat

Data získaná z mobilních telefonů pomocí Časové přímky je nutné po jejich předání upravit pro další zpracování do podoby vhodné pro analýzy a zároveň zajistit jejich dostatečnou anonymizaci agregací na čtvercovou síť. Většina kroků zpracování je provedena v softwaru ArcMap 10.3.

Po převedení dat do vrstvy shapefile je nutné data začistit na požadovaný časový úsek, tedy od 1. 10. 2015 do 14. 10. 2015. U dat předaných den po dni tento krok odpadá. Začištění je možné provést v atributové tabulce vrstvy, kdy vymažeme body přesahující časový úsek.

Dalším krokem je vymazání přebytečných či redundantních atributů pomocí vestavěné funkce softwaru ArcMap 10.3 Delete Field. V datech po úpravě zůstanou pouze atributy ID prvku, tvar prvku a časová značka.

Původní data z Časové přímky jsou v souřadnicovém systému WGS 84 (World Geodetic System 1984). Pro jednodušší kombinaci dat s mapovými podklady jsem data převedla do souřadnicového systému S-JTSK Krovak EastNorth (Systém jednotné trigonometrické sítě katastrální). K převodu dat z jednoho souřadnicového systému do jiného slouží v softwaru ArcMap 10.3 nástroj Project. Je k tomu potřeba znát převodní vztahy mezi jednotlivými systémy, některé vztahy jsou ve funkci dostupné defaultně. [2] Transformace probíhá ve dvou krocích. Prvním krokem je převedení referenčního elipsoidu WGS na elipsoid Besselův a druhým převedení Besselova elipsoidu na pravoúhlý souřadnicový systém S-JTSK. [2]

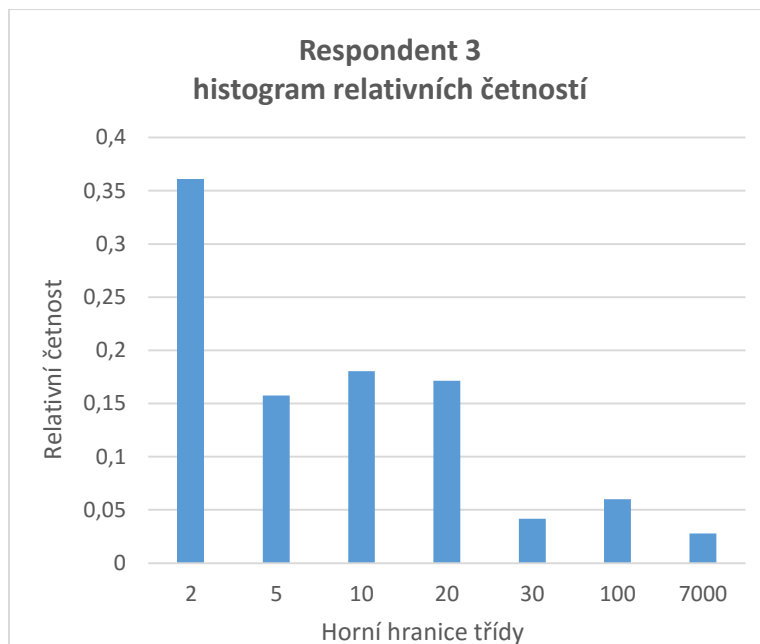
6 EDA a ESDA

Trasy naměřené respondenty jsem podrobila explorativní analýze dat. Zaměřila jsem se na aktivitu respondentů v prostoru, jak obecně tak v konkrétních čtvercích gridu. Ve čtvercové síti jsem pro každý čtverec získala hodnotu počtu linií nacházejících se v něm nebo čtvercem procházejících. Tyto četnosti jsem následně pomocí programu MS Excel a GeoDa vyhodnotila ze statistického hlediska. Pro každého respondenta jsem vypočetla aritmetický průměr četností ve čtvercích, medián, směrodatnou odchylku, koeficient špičatosti a šikmosti, maximum a počet čtverců, kterými prochází alespoň jedna linie. Z explorativní analýzy dat ESDA jsem zvolila percentilovou mapu, box plot mapu a také jednoduchý proporcionální kartogram.

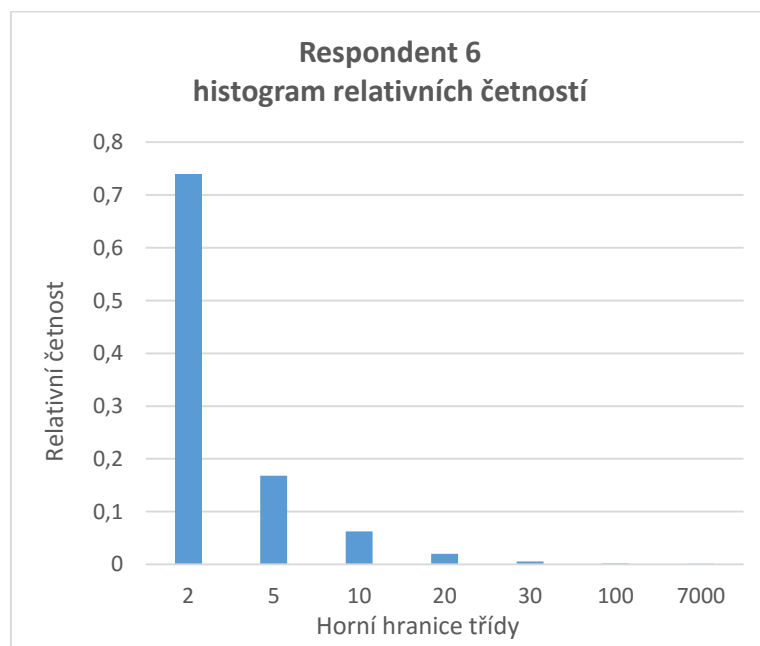
Tabulka 2. Statistika EDA nad daty respondentů

R	\bar{x}	\tilde{x}	σ	Špičatost	Šikmost	Maximum	Počet
1	16,57	2	245,83	742,34	27	6776	774
2	6,3	1	60,38	950,49	28,76	2235	2229
3	79,49	5	633,24	99,35	9,93	6601	216
4	3,84	1	27,72	621,53	23,43	762	910
5	3,9	2	25,27	638,67	23,64	675	851
6	3,19	1	27,48	755,32	55,86	1647	3862
7	13,79	2	79,7	374,91	17,03	2007	1109

Z aritmetického průměru lze vyzorovat mobilitu respondenta z hlediska aktivity. Respondenti setrvávající na stejných místech a projíždějících stále po stejných trasách mají aritmetický průměr vyšší, naopak aritmetický průměr respondentů, kteří často cestují, je velmi nízký (kolem 3 – 6). Medián, jako další střední hodnota opět určuje aktivitu respondenta ve vybraném území. Vyšší hodnoty ukazují na menší aktivitu. Šikmost a špičatost jsou u všech respondentů na vysokých kladných hodnotách. Statistický soubor má levostranné rozdělení, protože je šikmost větší než jedna a u většiny respondentů se pohybuje v hodnotách kolem 20. Špičatost je velmi vysoká u těch, kteří cestují více. Rozdíl hodnot špičatosti mezi nejaktivnějším a nejméně aktivním respondentem je 850. Hodnoty četností v souboru se tedy příliš neliší.



Obr. 1. Histogram relativních četností pasivního respondenta.

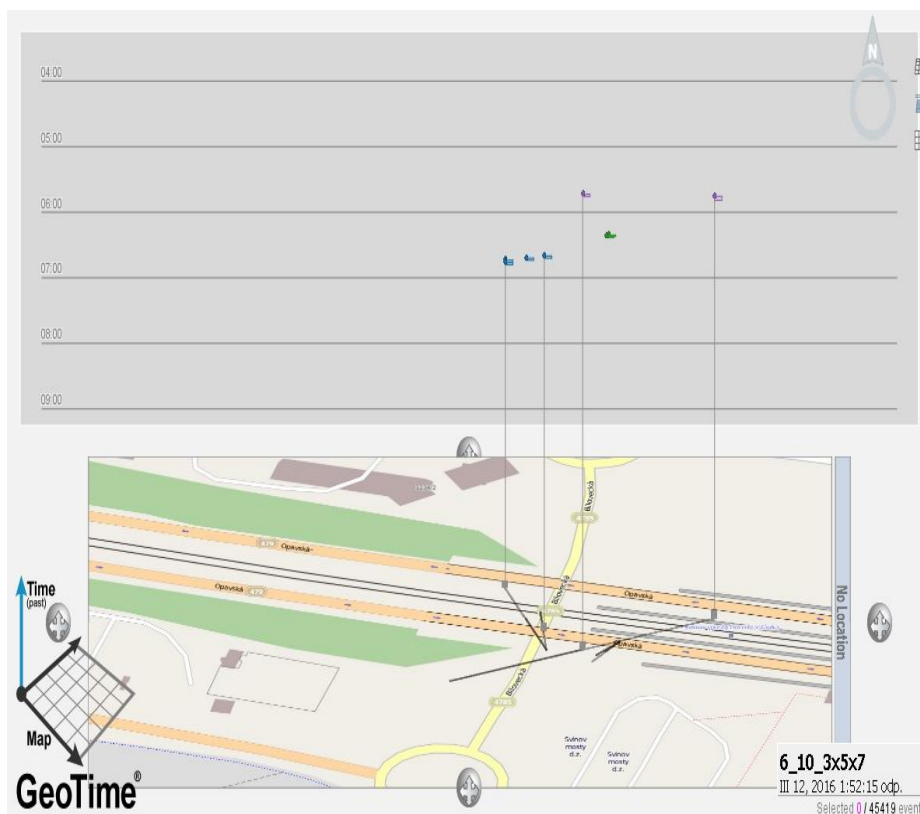


Obr. 2. Histogram relativních četností aktivního respondenta.

Z histogramů je patrné levostranné rozdělení obou statistických souborů, předchozí výpočty koeficientu šikmosti jsou tedy správné. Distribuce souborů je homogenní. Pro možnost srovnání jsou histogramy v relativních číslech. Osa X představuje počet linií ve čtverci, popisné číslo osy je vždy horní hranicích intervalu hodnot. Osa Y pak četnost těchto počtů v jednotlivých čtvercích. Nejčtenějšími intervaly jsou (0; 2>, kde se nachází čtverce, kterými prochází jen jedna či dvě linie. U pasivního respondenta je patrné plošší rozdělení souboru, naproti tomu aktivní respondent má rozdělení velmi špičaté.

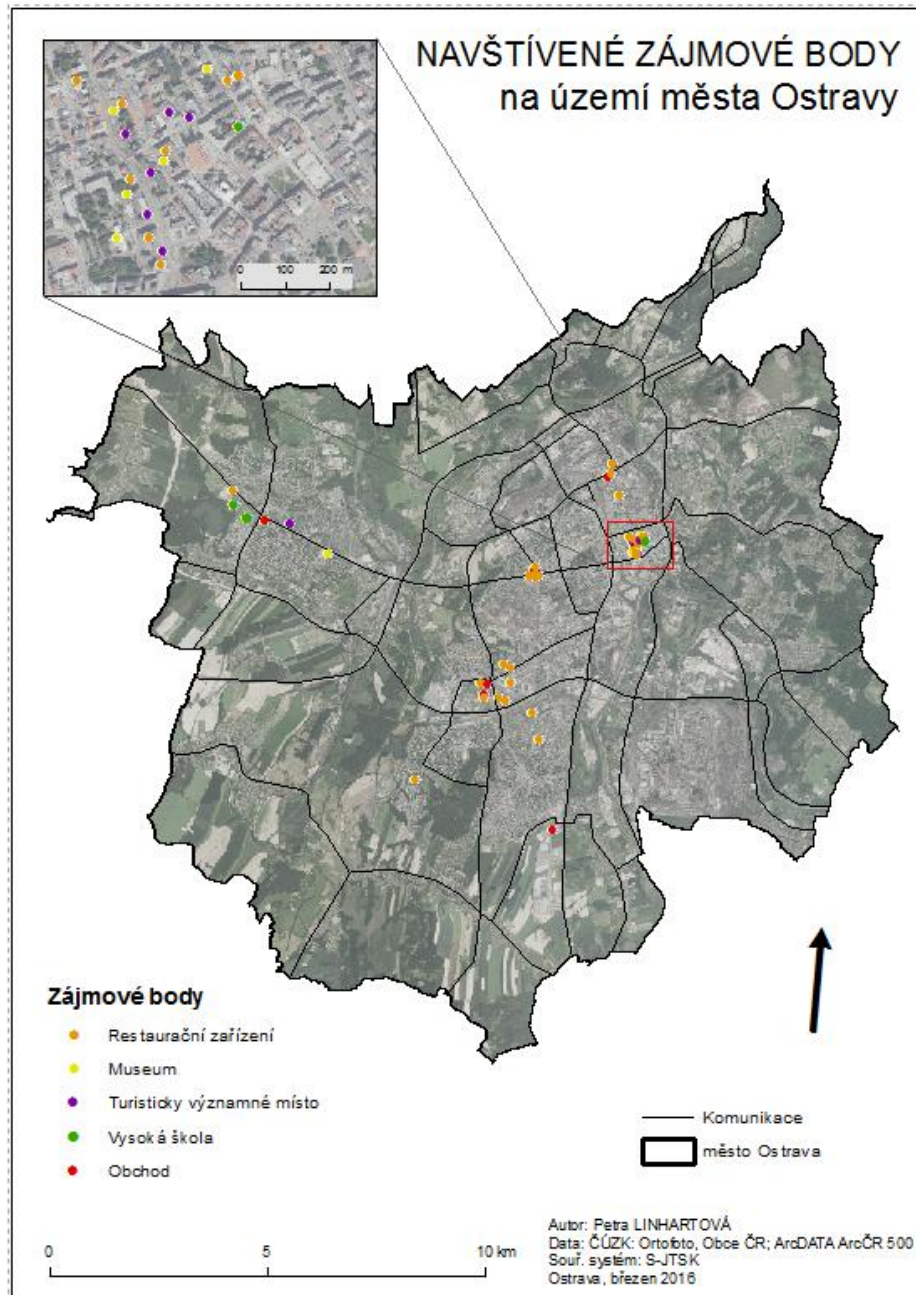
7 Časoprostorová analýza

V rámci software GeoTime jsem provedla časoprostorové analýzy hledající setkání v čase a prostoru a dále shluky dat mezi studenty. Nalezené shluky se nacházely nejvíce kolem Vysoké školy Báňské a částečně v okolí nádraží Ostrava Svinov. Měšší shluky byly nalezeny v centru města a také v ostravských Vítkovicích.



Obr. 3. Setkání respondentů v Porubě 6. 10. 2015.

8 Points of interests



Obr. 5. Mapa navštívených zájmových bodů na území města Ostravy.

Analýza proběhla nad čtveřcovou sítí každého respondenta zvlášť. U každého respondenta jsem provedla průnik čtveřcové sítě s bodovou vrstvou zájmových bodů Points of Interests. Od každého studenta jsem vzala body navštívené více než 10 krát a následně tyto body jednotlivých respondentů porovnála mezi sebou a vybrala 100 nejnavštěvovanějších bodů. Studenti velmi často navštěvují různá restaurační zařízení, ale také musea a turisticky významná místa jako jsou galerie či divadla.

Reference

1. Jakub Novák, Jana Temelová. Každodenní život a prostorová mobilita mladých pražanů: pilotní studie využití lokalizačních dat mobilních telefonů. *Sociologický časopis*, 2012, Vol. 48, No. 5: 911-938.). Praha 2012. ISBN.
2. Doc. Ing. Zdeněk Hrdina, CSc., Transformace souřadnic ze systému WGS-84 do systému S-JTSK,
<http://www.geospeleos.com/Mapovani/WGS84toSJTSK/WGS84toSJTSK.htm>.