Univerzita Palackého v Olomouci

Přírodovědecká fakulta

Katedra geoinformatiky



Ivan MATĚJČEK

REALIZACE METOD KLASIFIKACE KVANTITATIVNÍCH DAT V GIS

magisterská práce

Vedoucí práce: Mgr. Zuzana Němcová Olomouc 2010

Prohlašuji, že jsem zadanou magisterskou práci řešil sám a že jsem uvedl veškerou použitou literaturu.

Olomouc, 30. dubna 2010

.....

Podpis

Děkuji Mgr. Zuzaně Němcové za vedení diplomové práce. Dále děkuji doc. RNDr. Jaromíru Kaňokovi, Csc., Ing. Zdeně Dobešové, Ph.D. a Mgr. Pavlu Tučkovi, Ph.D. za cenné rady a připomínky.

<u>Obsah</u>

1. Úvod	8
2. Cíle práce	9
3. Metody a postup zpracování	.11
3.1. Studium odborné literatury	.11
3.2. Výběr SW a dat	.11
3.3. Hodnocení funkcionality SW	12
3.3.1. Jednoduchá binární metoda	12
3.3.2. Verbální hodnocení	12
3.4. Implementace metod do ArcGIS 9.3	12
3.4.1. Microsoft Visual Basic 6.3	. 13
3.5. Prezentace práce	14
4. Data	. 15
5. Klasifikační metody	. 18
5.1. Klasifikace s konstantní velikostí třídy	. 18
5.2. Klasifikace s proměnnou velikostí třídy	. 19
5.2.1. Metoda kvantilů	. 19
5.2.2. Metoda směrodatné odchylky	20
5.2.3. Metoda geometrické klasifikace	20
5.2.4. Armstrong-Xiaovo kritérium	
5.2.5. Caspall-Jenksovo kritérium	22
5.2.6. Ruční stanovení tříd	22
5.2.7. Metoda mediánová kumulativní klasifikace	22
6. Testované SW	24
6.1. Open source softwary	24
6.1.1. GRASS 6.4.0	25
6.1.2. Quantum GIS 1.0.1	25
6.1.3. uDig 1.1.0-SC3	26
6.2. Komerční programy	

6.2.1. ArcGIS 9.3
6.2.2. AutoCAD Map 3D 2009
6.2.3. Geomedia Professional 6.0
7. Hodnocení SW
7.1. Tvorba kartogramů a kartodiagramů29
7.2. Klasifikační metody
8. Vhodnost použití metod
8.1. Metody vhodné pro logaritmické rozdělení
8.2. Metody vhodné pro normální rozdělení
9. Implementace do ArcGIS 9.3
9.1. Vytvoření Toolbaru
9.2. Vytvoření formuláře
9.3. Tvorba skriptu
9.3.1. Definice proměnných
9.3.2. Inicializace formuláře
9.3.3. Procedura cbLayer_Change
9.3.4. Procedura cbAtribute_Change
9.3.5. Procedura cbMethod_change
9.3.6. Procedury tlačítek
9.3.7. Procedura CreateClass
9.3.8. Procedura CreateMap44
9.3.9. Procedura MedianClass
9.4. Vhodnost použití metody47
10. Diskuse
11. Závěr
12. Seznam použitých zdrojů
Summary
Přílohy

1. <u>Úvod</u>

Neustálé se zvyšující pronikání geoinformačních technologií do veřejného i soukromého sektoru má za následek to, že jsme nuceni přijímat a rozhodovat se podle jejich výsledků.

Tyto výsledky musí být srozumitelné pro širokou veřejnost, proto se nejčastěji setkáváme s tabulkami a grafickými výstupy, především mapami.

Mapy mají mnohem větší názornost a je v nich obsaženo více informací. To je důvod, proč jsou nejpoužívanější možností prezentace v geografickém informačním systému (GIS).

GIS zpracovává kvalitativní a kvantitativní data. Kvantitativní data se zobrazují v tzv. statistických mapách. Statistickými mapami se rozumí např. kartogramy a kartodiagramy.

Pro dosažení a ukázání problému nebo jeho řešení je v těchto mapách nutné použít správnou klasifikační metodu. Ta rozhoduje o tom, jak budou data tříděna do kategorií. Metod je velké množství, ale ne všechny ukazují to, co je potřebné zdůraznit. Tím se klasifikační metoda stává nejdůležitější při tvorbě statistické mapy.

GIS softwary se od sebe liší jak svým pracovním prostředím tak i funkcionalitou. To se týká i možností a množství nabídek při tvorbě statistických map.

Tato magisterská práce se snaží ukázat stávající možnosti klasifikace kvantitativních dat ve vybraných GIS produktech. A na příkladu jednoho produktu ukázat, že i uživatel si může doprogramovat další klasifikační metodu.

2. <u>Cíle práce</u>

Tato magisterská práce má čtyři cíle:

- Popsat stávají možnosti klasifikace kvantitativních dat
- Určit možnosti klasifikace v různých softwarech
- Prezentovat rozdíly při použití jednotlivých metod klasifikace
- Implementovat metodu do ArcGIS 9.3

Popis stávajících možností klasifikace je teoretická část této páce. Základem je se seznámit s různými klasifikačními metodami používaných v moderní kartografii u nás i ve světě pomocí odborné literatury a článků. Na základě získaných informací se dají hodnotit jednotlivé klasifikační funkce softwaru.

Funkcionalita je popsána pro šest softwarů, z nich jsou tři open source a tři komerční programy. Je hodnocena jednoduchým způsobem. V prvním kroku se zjistí, zda program dokáže vytvořit kartogramy a kartodiagramy. Za druhé se určí klasifikační metody, které daný software používá. Počet metod je měřítkem míry jeho funkcionality klasifikace kvantitativních dat.

Při testování softwaru vznikly i mapové výstupy. Na jejich základě je prezentována vhodnost použití klasifikačních metod pro různá data. Jako ukázkové slouží dva datasety. Jeden je v normálním a druhý v logaritmickém rozdělení.

Implementace metody v tomto případě znamená naprogramování daného algoritmu do softwaru firmy ESRI ArcGIS 9.3. Touto implementací se rozšíří funkcionalita tohoto programu. Programovacím jazykem je Microsoft Visual Basic 6.3. Výsledný skript je do programu nahrán jako nový toolbar. Zdrojové kódy jsou dobře okomentované v anglickém jazyce.

Pro tuto magisterskou práci je vytvořena internetová stránka a jednostránkové resumé v anglickém jazyce. Toolbar a zdrojové kódy skriptu

jsou k dispozici na CD-ROMu a na dané www stránce, kde ho bude možné stáhnout. Průvodní zpráva v tištěné verzi bude na Katedře geoinformatiky UP v Olomouci.

3. Metody a postup zpracování

3.1. Studium odborné literatury

Pro naplánování a určení postupu práce bylo nejdříve nutné nastudovat odbornou literaturu. Pro účely této magisterské práce se studovaly především kartografické a statistické publikace. Bylo čerpáno jak z české tak zahraniční literatury.

Další nutnou literaturou byly manuály a uživatelské příručky k jednotlivým softwarům. Hlavně k open source produktům, se kterými nebylo tolik předešlých zkušeností. V neposlední řadě byly nastudovány i návody k programování v MS Visual Basic 6.3 a jeho implementace do prostředí ArcGIS 9.3.

3.2. Výběr SW a dat

Po nastudování odborné literatury se musely vybrat softwary, u kterých bude testována jejich funkcionalita. Byly vybrány tři zástupci open source programů (GRASS, QuantumGIS, uDig) a stejný počet komerčních GIS softwarů (ArcGIS, AutocadMap 3D, Geomedia), které jsou dostupné na Katedře geoinformatiky Univerzity Palackého v Olomouci.

Když už byl vybraný software, začala se hledat vhodná data. Data se vybírala z datasetu Sčítání lidu, bytů a domů v roce 2001 od Českého statistického úřadu.

Po nainstalování vybraných open source produktů byly vyzkoušeny jejich možnosti klasifikace kvantitativních dat. A na jejich základech byly softwary zhodnoceny.

3.3. Hodnocení funkcionality SW

Když byly vybrány softwary a data, mohlo se přistoupit k hodnocení funkcionality. Byly využity dvě metody hodnocení. První byla jednoduchá binární metoda a druhá verbální hodnocení.

3.3.1. Jednoduchá binární metoda

Je to základní metoda, kde se hodnotí 1 nebo 0. To znamená, že danou klasifikační metodu lze resp. nelze v programu použít [3]. Z výsledků tohoto postupu je možné rychle určit jaké možnosti má daný software.

Jednoduchost však má i nevýhody. Jednou z nich je, že se programy nedají vzájemně porovnat. Porovnání softwarů je jedním z předpokládaných výstupů, aby byla tato podmínka splněna, metoda se rozšířila o součet "zásahů". Zásahem se myslí, že metoda je v softwaru dostupná (binárně 1). Čím byl větší součet, tím se program umístil lépe.

Tato metoda se použila při určení možnosti tvorby kartodiagramů a při dostupnosti klasifikačních metod.

3.3.2. Verbální hodnocení

Toto hodnocení je vyjádřeno slovním popisem, kdy lze slovně popsat výhody a nevýhody [14]. Je to častá metoda založená na přirozeném názoru hodnotitele. Spočívá ve výčtu kladů a záporů. Toto hodnocení má někdy charakter recenze [2].

Tato metoda doplňovala předchozí při hodnocení dostupnosti metod klasifikace v jednotlivých programech.

3.4. Implementace metod do ArcGIS 9.3

Pro implementaci metody byl vybrán software ArcGIS 9.3. Jeho hlavní předností je jeho celosvětové rozšíření. Tudíž nově vzniklý skript bude moci mít široké uplatnění.

Skript je psán v programovacím jazyku Microsoft Visual Basic 6.3. Pomocí tohoto jazyka byl vytvořen toolbar pro software ArcGIS 9.3. Software je popsán v kapitole 6.2.1.

3.4.1. Microsoft Visual Basic 6.3

Podle [8] je Microsoft Visual Basic 6 objektově orientovaný programovací jazyk. To znamená, že programátor může používat velké množství předdefinovaných objektů. Souhrnně se tyto objekty nazývají ovládací prvky (anglicky controls). Každý ovládací prvek má definovány své vlastnosti, metody a události [8].

Vlastnosti se dají definovat programovým kódem. Lze rozdělit do kategorií. Kategoriemi mohou být vzhled, chování, atd.[8]

Metody ovládacích prvků představují činnosti, které daný prvek může vykonávat, případně které mohou být vykonány na něm[8].

Události definované u všech ovládacích prvků zajišťují, že programování v Microsoft Visual Basic se počítá mezi událostmi řízené programování. U každého ovládacího prvku najdeme seznam událostí, které mohou při běhu programu vzniknout v přímém vztahu k tomuto prvku [8].



Obr. 1: Ukázka pracovního prostředí v Microsoft Visual Basic

3.5. Prezentace práce

Výsledky této práce lze nalézt na několika místech. Vytisknutá průvodní zpráva je k dispozici na Katedře geoinformatiky UP v Olomouci.

Elektronická podoba práce je dostupná na webových stránkách, které byly kvůli tomu vytvořené. Tyto stránky jsou pouze statické a dá se na nich mimo průvodní zprávy stáhnout i projekt s toolbarem pro ArcGIS 9.3.

4. <u>Data</u>

Pro testování softwarů se používala data z Českého statistického úřadu (ČSÚ) ze Sčítání lidu, domů a bytů k 1. březnu 2001 za Olomoucký kraj.

Data byla poskytnuta ve formátu Microsoft Excel (přípona .xls), proto bylo nutné vybrané atributy propojit s vektorovou vrstvou ve formátu shapefile (.shp) od ESRI. Jako vhodná vrstva byla vybrána polygonová vrstva obcí z ArcČR 2.0, ze které byly vybrány obce spadající pod Olomoucký kraj. K propojení bylo použito kódu obce.

Bylo určeno, že budou testovány dva datasety. Jeden v logaritmickém a druhý v normálním rozdělení.

Logaritmické rozdělení (obr. 2) je reprezentováno hustotou obyvatel. Jednotkou jsou obyvatelé na km². Hustota byla spočítána tak, že počet obyvatel v obci se vydělil rozlohou dané obce v km², která byla udána v ArcČR 2.0. Hodnotu hustoty nese atribut *hustota*.

Z tohoto datasetu vznikaly kartogramy, což je podle [13] jednoduchá tematická mapa, kde pro každý areál implementujeme jednu nebo více relativních hodnot vztažených k ploše. Metoda kartogramů je nejpoužívanější statistickou metodou v kartografii [14].

Logaritmické rozdělení bylo ověřeno tak, že data byla zlogaritmována a na tyto data byl použit Shapiro-Wilkův test normality. Hodnota tohoto testu byla p-value = 3,691e-06. Hodnota je menší než 0,05. Tudíž se jedná o logaritmické rozdělení.



Obr. 2: Graf četností hodnot pro hustotu (vytvořeno v programu R)

Normální rozdělení (obr. 3), neboli Gaussovo, je základní. Z něho vychází většina matematických postupů. Toto rozdělení reprezentuje podíl žen na celkovém počtu obyvatel. Hodnota byla spočítána tak, že se počet žen vydělit celkovým počtem obyvatel v dané obci. Výsledek obsahuje atribut *zeny_celke*.

V tomto případě se jedná o nepravý kartogram, protože jev není přepočítán na plochu [19]. K nepravému kartogramu se přistoupilo z důvodu nemožnosti z daných dat vytvořit kartogram pravý tak, aby data byla v normálním rozdělení.

Normalita rozdělení byla opět otestována v programu R pomocí Shapiro– Wilkova testu (p-value = 3,987e-14).



Obr. 3: *Graf četností hodnot pro podíl žen na celkovém počtu obyv. (vytvořeno v programu R)*

Výsledný shapefile se jmenuje *OLkraj.shp*. Obsahuje všechny výsledné atributy a atributy, ze kterých byly spočítány.

Z důvodu, že daná data jsou již zaregistrována v metadatovém sytému katedry, není nutné je tedy znovu registrovat. Výsledné hodnoty se dají kdykoli z dostupných dat znovu dopočítat.

5. Klasifikační metody

Klasifikační metody mají v kartografii široké uplatnění. Používají se při tvorbě jednoduchých i složitých tématických map, v kterých je využívána metoda kartogramů. Další aplikací může být metoda barevné hypsometrie při znázorňování terénu.

Klasifikace přerozděluje posloupnost *N* hodnot $h = \{h_1, ..., h_N\}$ do *k* tříd, kde platí $k \le N$. Výsledkem klasifikace je posloupnost hranic tříd. To znamená, že každé hodnotě h_i lze přiřadit pořadové číslo třídy, do které spadá. Cílem klasifikace je, aby hodnoty prvků h_i uvnitř jedné třídy byly co nejpodobnější. Neměly by vznikat prázdné třídy nebo naopak třídy, kde se četnost prvků blíží k celkovému počtu prvků posloupnosti *N* [1].

Rozložení hodnot h_i ve vstupním souboru hraje důležitou roli. Pokud odpovídá normálnímu, popř. konstantnímu rozložení. Je možné použít základní klasifikační metody (např. klasifikaci s konstantní velikosti tříd). Pokud soubor nemá toto rozložení, je nutné použít speciální klasifikační postupy, tj. klasifikaci s proměnnou velikostí tříd[1].

Před použitím klasifikačního postupu je vhodné zjistit rozložení hodnot h_i ve vstupním souboru a otestovat, zda tyto hodnoty představují výběr s normálním rozdělením[1].

Pro testování lze použít například **Shapiro-Wilkův test normality výběru** [1].

Rozlišujeme klasifikaci s konstantní velikostí třídy a klasifikaci s proměnnou velikostí třídy.

5.1. Klasifikace s konstantní velikostí třídy

Podle [1] se tato metoda používá, když se rozdělení dat blíží normálnímu nebo konstantnímu. Vstupní údaj představuje počet tříd k, do kterých bude posloupnost N hodnot rozdělena.

Velikost třídy se určí z následujícího vztahu:

$$\Delta I = \frac{h_{\max} - h_{\min}}{k} = konst.$$
(5.1)

Hodnota h_{max} je maximální hodnota jevu a h_{min} představuje naopak minimální hodnotu. Počáteční hranici *Ij* stanovíme jako:

$$I_j = h \min + j \Delta I , \qquad (5.2)$$

kde j = 1, ..., k - 1

Podle [6] přepočtem vzorce 5.1 získáme:

$$h_{\min} + k\Delta I = h_{\max} \tag{5.3}$$

V softwarech se tato metoda většinou nazývá **Equal Interval**. Obdobou je metoda **Defined Interval**, kde uživatel nezadává počet tříd, ale přímo určí velikost třídy.

5.2. Klasifikace s proměnnou velikostí třídy

Metoda je používána když je rozložení hodnot nerovnoměrné. Jak už z názvu plyne, budou velikosti tříd různé. Do této skupiny paří následující metody.

Żádná z těchto metod není univerzálně použitelná. Jejich použití závisí na typu dat. Nevhodně použitá metoda bude mít za následek, že vzniknou prázdné třídy nebo četnost v třídách bude velmi nerovnoměrná.

5.2.1. Metoda kvantilů

Podle [1] a [6] je to klasifikační metoda, při které se uvnitř každé třídy nachází předem známý počet prvků. (pokud N/k je celé číslo). Většinou se používá při konstantním rozložení.

Podle počtu tříd tvoří hranice kvartily (čtyři třídy), quintily (pět tříd) atd. Hraniční hodnotu Q_i lze určit z následujícího vztahu:

$$Q_j = (N+1)\frac{J}{q} \tag{5.4}$$

Kde *N* je celkový počet prvků, *q* prezentuje kvantily (kvartily, quintily,...) a *j* představuje o kolikátou hranici se jedná. Např. pro q = 5, j = 1,2,3,4.

Nevýhodou je, že do jedné třídy mohou být sloučeny numericky velmi rozdílné hodnoty, když není splněna podmínka konstantního rozložení.

V softwarech tuto metodu najdeme pod názvem Quantile nebo Quantiles.

5.2.2. Metoda směrodatné odchylky

Podle [1] a [6] je metoda založena na analýze směrodatné odchylky σ souboru s normálním rozložením hodnot. Nejprve je vypočtena střední hodnota prvku \overline{h} a následně směrodatná odchylka σ .

Hranice intervalů bývají počítány symetricky na obě strany od střední hodnoty přičtením nebo odečtením *c* násobku σ (*c* = 0.5, 0.25, 0.125) dokud nejsou všechna data umístěna ve třídách. Hodnoty přesahující ±3 σ jsou sloučeny do jedné třídy, tudíž okrajové třídy mají jinou velikost než třídy vnitřní. Pro hranice tříd platí vztah:

$$Ij = \overline{h} \pm cj \sum_{j=1}^{k} \sigma$$
(5.5)

Tato metoda přiřazuje prvky do tříd podle toho, o kolik se liší od hodnoty \overline{h} . Metoda je vhodná pro data v normálním rozdělení.

Ve většině GIS tato metoda nese pojmenování **Standard Deviation**, popřípadě zkráceně **StdDev**.

5.2.3. Metoda geometrické klasifikace

Geometrická klasifikace se podle [6] počítá následovně. Každý následující člen je odvozen z předešlého tak, že se vynásobí konstantou *C*, což je poměr tříd.

Nejdříve se musí určit logaritmus nejvyšší a nejnižší hodnoty. Z těchto dvou hodnot se spočítá konstanta *C* jako:

$$C = \log h_{\max} - \log h_{\min} / n \tag{5.6}$$

Jednotlivé logaritmy hranic tříd se určí podle vzorce:

 $\log h_{\max} - C = \log 2$. nejvyšší hraniční hodnoty (5.7)

 $\log 2$. nejvyšší hraniční hodnota – $C = \log 3$. nejvyšší hraniční hodnota,(5.8)

toto opakujeme tolikrát, kolik je zvoleno tříd, např. pro 5 tříd postup zopakujeme 4krát.

Číslo hraniční hodnoty získáme tak, že výsledný logaritmus bude mocninou při základu 10. Např. log $h_{\text{max}} - C = 2$, z toho plyne, že hraniční hodnota bude 10^2 .

5.2.4. Armstrong-Xiaovo kritérium

Podle [1] kritérium minimalizuje sumu kvadrátů oprav od aritmetického průměru \overline{h}_j v jednotlivých třídách dělenou sumou kvadrátů oprav aritmetického průměru \overline{h} všech hodnot. Lze ho vyjádřit:

$$AX = 1 - \frac{\sum_{j=1}^{k} \sum_{i=1}^{N} (h_{ij} - \overline{h}_{j})^{2}}{\sum_{i=1}^{N} (h_{i} - \overline{h})^{2}}$$
(5.9)

Hodnoty h_i , kde i = 1, 2, ..., N, představují měřené údaje, k počet tříd, \overline{h}_j aritmetický průměr ve třídě j, N_j počet hodnot ve třídě j. Jmenovatel představuje konstantu, proto je nezávislý na volbě počtu tříd.

Detailně popsaný algoritmus tvorby Armstrong-Xiaova kritéria naleznete v knize T. Bayera: Algoritmy v digitální kartografii.

Kritérium je také někdy označováno jako Goodness of Variance Fit. V GIS programech se nejčastěji vyskytuje pod názvem **Jenks natural breaks**, u kterého je vhodné používat heuristický shlukovací algoritmus k-means. Tím pádem neposkytuje exaktní řešení, ale na druhou stranu nevytváří řešení špatná.

5.2.5. Caspall-Jenksovo kritérium

Podle [1] je podobné metodě Armstrong-Xiaovu kritériu. Jiný je způsob výpočtu. Chybový člen představuje chybu absolutních hodnot oprav od aritmetického průměru \overline{h}_j v jednotlivých třídách je dělenou sumou absolutních hodnot oprav aritmetického průměru \overline{h} všech hodnot. Není používána euklidovská metrika, ale manhattanská. Caspall-Jenksovo kritérium lze vyjádřit:

$$CJ = 1 - \frac{\sum_{j=1}^{k} \sum_{i=1}^{N} \left| h_{ij} - \overline{h}_{j} \right|}{\sum_{i=1}^{N} \left| h_{i} - \overline{h} \right|}$$
(5.10)

V tomto kritériu se dají použít i váhy. Váha představuje plochu uzavřených oblastí. Proto ho lze použít pouze u dat reprezentujících nějaký jev vztažený na plochu. Tímto se stává vhodným doplňkem tematických map [1].

Toto kritérium chybí ve všech softwarech. Výjimkou není ani ArcGIS 9.3, avšak její implementace je velmi obtížná z důvodu využívání heuristiky v algoritmu.

5.2.6. Ruční stanovení tříd

Podle [1] v některých případech jsou automatizované algoritmy nevhodné. V tom případě je potřeba nastavit intervaly ručně. Jedná se hlavně o případy, kdy data mají speciální rozdělení, či je potřeba znázornit zvláštní charakteristiky zobrazovaných dat. Případně si uživatel sám spočítal hranice tříd a nastaví je. V programech se vyskytuje většinou jako **Manual**.

5.2.7. Metoda mediánová kumulativní klasifikace

Mediánová kumulativní klasifiakce je nově navržená metoda, která vychází z předpokladu, že medián \tilde{h} dělí soubor všech hodnot *h* na dvě stejně velké skupiny. Pro každou z těchto dvou skupin jsou potom odvozeny velikosti

intervalů podle vztahu 5.11 pro hodnoty menší než \tilde{h} . A podle vztahu 5.12 pro hodnoty větší než \tilde{h} .

Tyto vztahy vychází z rovnice konstantních intervalů (aritmetická posloupnost). V tomto případě však členy nenarůstají konstantně, ale podle daných vztahů. Tím pádem velikost intervalu je závislá na počtu intervalů a pořadí členu.

$$a_n = a_{(n-1)} + (p - n + 1)d_1, \qquad (5.11)$$

kde *a* je hraniční hodnota intervalu, *p* je polovinou počtu klasifikačních tříd, *n* udává pozici hranice intervalu a d_1 je koeficient pro intervaly s hodnotami menšími než \tilde{h} . Vypočítaný podle vzorce 5.13

$$a_n = a_{(n-1)} + (n-1)d_2, \qquad (5.12)$$

kde jsou všechny veličiny shodné a d_2 je koeficient pro intervaly s hodnotami vyššími než \tilde{h} . Hodnota d_2 se spočítá podle vztahu 5.14.

$$d_1 = \frac{2(\tilde{h} - h_{\min})}{p^2 + p}$$
(5.13)

$$d_2 = \frac{2(h_{\max} - \tilde{h})}{p^2 + p}$$
(5.14)

Tato metoda člení data do sudého počtu tříd, aby byl zachován medián jako střed a na obě strany byl shodný počet tříd.

V této metodě můžeme místo mediánu použít jakoukoliv zvolenou hodnotu např. průměr.

6. Testované SW

Softwary, které jsou testovány, byly vybrány tak, aby polovina z nich byla open source a druhá polovina komerční programy.

Byla snaha, aby testované verze programů měly pokud možno nejnovější verze. Tato podmínka byla splněna. Softwary jsou nejnovější k srpnu roku 2009, kdy probíhalo jejich testování.

6.1. Open source softwary

První testovanou skupinou byly tzv.open source softwary. Jak jejich anglický název napovídá mají otevřený svůj zdrojový kód. Jsou přístupné všem uživatelům. Díky své "otevřené" licenci. Tudíž je může využívat kdokoli zcela zdarma.

Většina těchto softwarů jsou publikovány pod GNU GPL (General Public License). Což je podle [21] všeobecná veřejná licence. To znamená, že zdrojové kódy a programy jsou svobodně dostupné a všechny odvozené programy musí být opět vydávány pod touto licencí.

U nás nejsou tolik využívané hlavně z důvodu, že nemají žádnou garantovanou podporu, proto by byly pro většinu subjektů, kde s těmito programy musí pracovat i laici velkým rizikem.

Jako zástupci open source programů byly vybrány tři programy a to GRASS, Quantum GIS a uDig. Všechny tři se dají stáhnout na svých domovských stránkách nebo na http://www.osgeo.org.

OSGeo (The Open Source Geospatial Foundation) je nezisková organizace. Jejím cílem je podporovat rozvoj a rozšiřování otevřených geoprostorových technologií a dat domovskou stránkou organizace najdeme na http://www.osgeo.org/content/foundation/about.html.

6.1.1. <u>GRASS 6.4.0</u>

Asi nejznámějším open source softwarem je GRASS (Geographic Resources Analysis Support System). Základ vyvíjel USA-CERL (U.S. Army Construction Engineering Research Laboratories) v letech 1982 - 1995. V roce 1997 se na vývoji podílí GRASS Development Team, který sídlí na Baylor University in Texas [22].

GRASS je rastrový a vektorový desktop systém. Dále dokáže zpracovávat spektrální snímky, provádět konverze mezi rastrovými a vektorovými daty, má nástroje pro tvorbu map a další. To z tohoto modulárně řešeného GIS, který využívá více než 350 programů a nástrojů, dělá velkého konkurenta komerčních GIS [22].

GRASS je spustitelný na mnoha operačních systémech (Linux, MS Windows, Mac OS X a dalších). Umožňuje pracovat v grafickém rozhraní nebo zadáváním do příkazového řádku. Samozřejmostí je doprogramování celých modulů nebo skriptů. Jako programovací jazyk se nejběžněji používá C++ [22].

Nejnovější verzí je GRASS 6.4.0. Verze 6 je vyvíjena od začátku roku 2005, verze 6.4.0 se objevila v prosinci 2008. Novinkou v GRASS 6 je nový engine pro práci s 2D a 3D vektorovými vrstvami a využívá DBMS (Systém řízení databáze) na bázi SQL [22].

Více informací a možnost stažení softwaru je na domovské stránce http://grass.itc.it.

6.1.2. <u>Quantum GIS 1.0.1</u>

Projekt Quantum GIS (zkráceně QGIS) se rozeběhl v květnu roku 2002. Začalo to nápadem Garyho Shermana, který hledal prohlížečku geodat pro Linux, která by byla rychlá a podporovala širokou škálu datových úložišť [17].

Časem se z QGIS začal stávat plnohodnotný desktop GIS, který běží na platformách Linux, MS Windows, Mac OS X. Je psán v jazyce C++, grafické

rozhraní je postavené na knihovně Qt. Zásuvné moduly jsou tvořeny v C++ nebo v Pythonu [17].

Díky jeho přehlednému a jednoduchému grafickému rozhraní, kde se dá pracovat jak s rastrovými tak vektorovými vrstvami, je využíván mnoha uživateli, kteří zobrazují a upravují data [17].

Zajímavostí je, že tento program obsahuje modul, který zpřístupňuje funkce GRASSu. Tím pádem se může stát jeho nadstavbou.

Existuje i česká lokalizovaná verze. Domovskou stránkou tohoto softwaru je http://qgis.org.

6.1.3. <u>uDig 1.1.0-SC3</u>

Program uDig je uvolňován pod licencí GNU Lesser General Public Licence. Projekt je sponzorován kanadskou společností Refractions Research, která se značnou měrou podílí i na jeho vývoji [26].

UDig je spustitelný na platformách MS Windows, Linux i Mac OS X. Je to Java aplikace založená na platformě Eclipse "Rich Client Platform" (RCP). Tvorba skriptů pro tento software je možná pomocí jazyka Java a vkládají se jako RCP plug-in [26].

Silnou stránkou tohoto programu je možnost čtení a zápisů Shapefile, PostGIS, DB 2, Oracle a ArcSDE. Dále jsou kompletně integrovány standardy OGC WMS (Web Map Services) a WFS (Web Feature Services) [25].

Program i jeho aplikace pro vývoje je možné stáhnout na stránce http://udig.refractions.net.

6.2. Komerční programy

Na rozdíl od open source programů tyto mají placené licence. To limituje možnost výběru programů pouze na ty, co jsou dostupné na Katedře geoinformatiky UP v Olomouci.

V praxi se s nimi však setkáváme častěji, protože jsou velmi dobře propagovány a mají silnou technickou podporu. Nespornou výhodou jsou i kurzy a školení, které firmy dodávající software nabízejí.

6.2.1. <u>ArcGIS 9.3</u>

Nejrozšířenějším komerčním softwarem je ArcGIS. Vytvořen firmou ESRI sídlící v Redlands v Kalifonii.

Je to GIS software který se prodává ve třech licencích [18]:

<u>ArcView</u> poskytuje rozsáhlé nástroje pro tvorbu map a získávání informací z map a jednoduché nástroje pro editaci a prostorové operace.

<u>ArcEditor</u> má plnou funkcionalitu ArcView a navíc rozšířené editační možnosti pro geodatabáze a shapefile.

<u>ArcInfo</u> obohacuje funkcionalitu obou předchozích produktů o rozšířené prostorové operace. Také umožňuje využít stávající aplikace pro ArcInfo Workstation.

Jednotná architektura umožňuje, že uživatelé pracující s kteroukoli z licencí, mohou sdílet výsledky své práce s ostatními uživateli. Mapy, data, symboly, mapové vrstvy, uživatelské nástroje a rozhraní, výstupní sestavy, metadata atd. mohou být vzájemně sdíleny a vyměňovány mezi všemi třemi produkty. Použití jednotné architektury přináší uživatelům i tu výhodu, že k ovládání kterékoliv aplikace ArcGIS Desktop se stačí naučit ovládat jednotné uživatelské rozhraní [18].

Funkcionalitu všech těchto produktů lze rozšířit přidáním nejrůznějších nadstaveb ArcGIS Desktop, vyvinutých ESRI a dalšími organizacemi. Mimoto si uživatelé mohou s využitím ArcObjects, knihovny softwarových komponent ArcGIS, vyvinout vlastní nadstavby ArcGIS Desktop. Tyto uživatelské nadstavby a nástroje mohou uživatelé vyvíjet ve standardních programovacích rozhraních Windows, jako je Microsoft Visual Basic, .NET, Java a Visual C++ [18].

6.2.2. AutoCAD Map 3D 2009

AutoCAD Map 3D je vytvořen nad základní aplikací AutoCAD od firmy Autodesk, která je zaměřena především na tvorbu CAD programů. Podle [4] jde tudíž o propojení funkcí CAD a GIS.

Software je postaven na platformě AutoCAD 2009. To znamená, že obsahuje všechny jeho nástroje. A k nim jsou přidávány klasické nástroje GIS jako například tvorba obalových zón, síťové analýzy atd. [4].

Je určen všem uživatelům GIS. Od správců systému až po obyčejné uživatele GIS. Uživatelé si musí přivyknout trochu jiné práci s vrstvami než v klasických GIS softwarech [4].

Program umožňuje integrovat mnoho typů a formátů dat. Dokáže pracovat jak s vektorovými tak i rastrovými daty. Z databází využívá Oracle Spatial, ArcSDE a další. Umí pracovat i s webovými službami WMS a WFS [4].

6.2.3. Geomedia Professional 6.0

Geomedia Professional je technologie od firmy Intergraph z USA. Geomedia je univerzální GIS klient, který umožňuje pracovat s řadou formátů od různých výrobců (Shapefile, dwg od Autodesku a další). Samozřejmostí je i podpora otevřených standardů jako WMS [23].

Umožňuje přenášet data z různých databází (Oracle Spatial, MS Access, SQL a mnoho dalších) a pracovat s několika databázemi najednou [23].

Tento software má široké využití v mnoha oborech. Hlavní jeho silou jsou prostředky pro navrhování a správu inženýrských sítí [23].

Vytváření vlastních aplikací a nadstaveb je možné pomocí OLE nástrojů a jazyků MS Visual Basic, Visual Basic .NET, C++. Software je stavěn pro platformu MS Windows [23].

7. Hodnocení SW

Pro hodnocení byly použity metody popsané v kapitole 3.3.1 resp. 3.3.2. To byly jednoduchá binární metoda a verbální hodnocení. Možnosti jednotlivých softwarů byly testovány na zkušebních datech.

Nejdříve bylo zjištěno, zda se v programu dají tvořit kartogramy a kartodiagramy. V dalším kroku se zjišťovalo množství klasifikačních metod.

7.1. Tvorba kartogramů a kartodiagramů

Než se začaly testovat samotné klasifikační metody bylo potřeba zjistit, zda se dají vůbec v daném softwaru tvořit kartogramy a kartodiagramy, při jejich tvorbě se klasifikace využívá.

To bylo zjišťováno pomocí dokumentace k programům a vlastní prací v nich. Výsledky jsou zaznamenány v tabulce 1.

Software	kartogram	kartodiagram
GRASS 6.4.0	1	1
QGIS 1.0.1	1	0
Udig 1.1.0 -SC3	1	0
ArcGIS 9.3	1	1
AutoCAD Map 3D 2009	1	0
Geomedia 6.0	1	0

Tabulka 1: Možnost tvorby kartogramů a kartodiagramů

V tabulce 1 můžeme vidět, že kartogramy je možné tvořit ve všech testovaných softwarech. Kartodiagramy lze tvořit pouze ve dvou a to v komerčním ArcGIS 9.3 a GRASS 6.4.0.

V některých dalších programech (např. AutoCAD Map 3D 2009) je možné diagramy vytvořit jiným způsobem, ale nemají přímo zakomponovaný nástroj tvorby kartodiagramů.

7.2. Klasifikační metody

Klasifikační metody jsou shodné jak pro tvorbu kartogramů tak kartodiagramů (jsou-li dostupné). Zkoušení funkcí probíhalo stejně jako u testování tvorby kartogramů a kartodiagramů.

V tabulce 2 jsou uvedeny všechny metody, které jsou popsány v kapitole 5. V jednotlivých softwarech se jejich názvy liší. Proto bylo potřeba všechny klasifikace vyzkoušet a porovnat s ostatními, aby se určilo, zda je jejich výpočtový algoritmus shodný. Porovnávání probíhalo vizuálně a podle hraničních hodnot tříd s přesností na jedno desetinné číslo. Jedna desetina kvůli programům, které hranici intervalů takto zaokrouhlovaly a nastavení se nedalo změnit (např. uDig).

Software	Konstantní intervaly	Definované intervaly	Kvantily
GRASS 6.4.0	1	0	1
QGIS 1.0.1	1	0	1
uDig 1.1.0 -SC3	1	0	1
ArcGIS 9.3	1	1	1
AutoCAD Map 3D 2009	1	0	1
Geomedia 6.0	1	1	0

Tabulka 2: Klasifikační metody při tvorbě kartogramů a kartodiagramů

Software	Směrodatná odchylka	Geometrická klas.	Armstrong-Xiaovo krit.
GRASS 6.4.0	1	0	1
QGIS 1.0.1	0	0	0
uDig 1.1.0 -SC3	0	0	0
ArcGIS 9.3	1	1	1
AutoCAD Map 3D 2009	1	0	1
Geomedia 6.0	1	0	0

Software	Caspal-Jenksovo krit.	Ruční stanovení	Mediánová kum. klas.
GRASS 6.4.0	0	1	0
QGIS 1.0.1	0	1	0
uDig 1.1.0 -SC3	0	1	0
ArcGIS 9.3	0	1	0
AutoCAD Map 3D 2009	0	1	0
Geomedia 6.0	0	1	0

Pro větší přehlednost byla vytvořena tabulka 3, která ukazuje konečné skóre jednotlivých softwarů. Skóre bylo spočítáno jako součet dostupných metod v každém programu. Na základě skóre bylo určeno pořadí softwarů.

Software	skóre	pořadí
GRASS 6.4.0	5	2 3.
QGIS 1.0.1	3	5 6.
uDig 1.1.0 -SC3	3	5 6.
ArcGIS 9.3	7	1.
AutoCad Map 3D 2009	5	2 3.
Geomedia 6.0	4	4.

Tabulka 3: Skóre a pořadí programů

Z tabulek 2 a 3 je vidět, že nejvíce metod nabízí ArcGIS 9.3 se sedmi klasifikačními metodami z devíti. Dalším softwarem byl open source GRASS 6.4.0 s pěti možnostmi klasifikace a společně s ním produkt firmy Autodesk AutoCAD Map 3D 2009. V ostatních programech byla dostupná méně než polovina klasifikačních metod.

Software ArcGIS má nejvíce klasifikačních metod a je v něm možnost tvořit jak kartogramy tak kartodiagramy. Díky jeho propracovanému systému tvorby map to z něj dělá nejlepší nástroj pro tvorbu statistických map, který byl testován. Nevýhodou je samozřejmě to, že je komerční. To znamená, že jeho pořizovací náklady nejsou zrovna malé.

Jako jeho open source alternativa vystupuje program GRASS. Má obdobné možnosti a to nejen v tvorbě statistických map.

Pro obyčejného uživatele se jako lepší open source alternativou jeví program QGIS a uDig, které mají velmi podobné grafické rozhraní s komerčními programy a jednoduše se do nich dají nahrávat různé datové formáty. Jednoduchost se bohužel projevuje i v menším počtu metod klasifikace i prováděných analýz.

Program QGIS může díky svojí nadstavbě GRASS využívat i funkce tohoto programu.

AutoCAD Map 3D ukazuje, že firma Autodesk se opravdu snaží proniknout do geoinformačních technologií. Tento software má mnoho funkcí klasického GIS a z pohledu testování nabízí mnoho klasifikačních metod. Jeho slabinou je pro uživatel zvyklé na prostředí GIS trochu jiné rozhraní. Nutno říci, že uživatelé tohoto programu jsou většinou projektanti a návrháři inženýrských sítí, kteří jsou v CAD systémech zvyklí pracovat.

Posledním zástupcem komerčních softwarů byl Geomedia Professional 6.0. Ten zaostal za očekáváním pouze s pěti klasifikačními metodami. Intergraph vytváří řešení pro mnoho oborů. Možná to zapříčinilo, že tato prezentační část programu není plně doceněna. Výhodou tohoto programu je, že si vytváří vlastní sklad dat a původní data nejsou měněna.

Každý z testovaných programů má své výhody a nevýhody. Toto jsou jen osobní zkušenosti a postřehy, s kterými nemusí každý souhlasit. Je to pouze taková malá recenze testovaných programů z mého pohledu.

8. Vhodnost použití metod

Cílem bylo určit vhodnost z hlediska charakteru dat, proto byly testovány dvoje data. Jedny v normálním a druhé v logaritmickém rozdělení.

Vhodnost využití metod byla určena vizuálním zhodnocením podle map vytvořených jednotlivými metodami. Mapy jsou k práci vloženy jako přílohy 2-9. Tento způsob hodnocení je subjektivní, proto závěry jednotlivých lidí nemusí být shodné.

Z posuzování byly vyloučeny "nestatistické" metody. Tím se rozumí metody, které nemají matematicky přesně definované hranice. Těmito metodami jsou ruční stanovení tříd a definované intervaly.

8.1. Metody vhodné pro logaritmické rozdělení

Klasifikační metody vychází z předpokladu, že data budou v normálním rozdělení. Často se stává, že tento předpoklad není naplněn. Nezkušený nebo neznalý uživatel si této skutečnosti nemusí všimnout a tím pádem tvoří nevypovídající a matoucí výstupy. Z tohoto důvodu byla testována i data, která nemají normální rozdělení.

Vizuální hodnocení bylo provedeno podle map v příloze 2, 3, 4, a 5. Výsledkem bylo, že nejvhodnější metodou, je metoda kvantilů.

Z klasifikace podle kvantilů (příloha 3) je patrná velká hustota obyvatel v okolí měst Olomouc, Přerov a Prostějov a pak podél hlavního silničního tahu na Jeseník. Periferní oblasti jsou většinou slabě osídlené.

Geometrická klasifikace (příloha 4) více zvýrazňuje jádrové oblasti. To znamená, že v prvním intervalu je více obcí než v prvním případě a naopak v posledním je méně. V této metodě zůstává pořád zachován trend osídlení u velkých měst a podél hlavní komunikace. Klasifikace podle Armstrong-Xiaova pravidla (příloha 5), jestě více zvýrazňuje hustě osídlené obce. Zde je už mnoho obcí v prvních třech intervalech.

Nejhorší metodou v tomto případě je metoda konstantních intervalů (příloha 2). Při této metodě je naprostá většina hodnot v prvním intervalu a v ostatních je vždy minimum hodnot.

8.2. Metody vhodné pro normální rozdělení

Pro normální rozdělení jsou vytvořeny všechny klasifikační algoritmy. Avšak ani zde nelze použít vždy všechny metody a je potřeba je pečlivě vybírat. Z testovaných dat hustoty obyvatel byly vytvořeny mapové výstupy, které jsou v přílohách 6, 7, 8, 9.

Vizuálním hodnocením byly jako nejvhodnější vybrány následující metody. Nejlepších výsledků bylo dosaženo metodou kvantilů (příloha 7).

Ještě vypovídající informace vznikly z metod geometrické klasifikace (příloha 8) a Armstrong-Xiaova kritéria (příloha 9). V těchto případech je už málo hodnot v krajních intervalech. Proto by bylo vhodné použití pro znázornění minimálních a maximálních hodnot.

Opět jako nejméně vhodná metoda se jevila metoda konstantních intervalů (příloha 6).

9. Implementace do ArcGIS 9.3

Posledním úkolem bylo implementovat novou metodu, kterou byla mediánová kumulativní klasifikace, do programu ArcGIS 9.3 pomocí skriptu. Skript byl psán v jazyce Microsoft Visual Basic 6.3.

Implementace probíhala v několika etapách, které jsou popsány níže.

9.1. Vytvoření Toolbaru

Nejdříve byl vytvořen toolbar nazvaný Classificator (obr. 4), v kterém je jedno tlačítko (nazvané také Classificator), které vyvolá dialogové okno (formulář), v kterém si uživatel volí parametry klasifikace.



Obr. 4: Náhled toolbaru Classificator

Zdrojový kód tlačítka Classificator je následující:

9.2. Vytvoření formuláře

Po té se přistoupilo k programování formuláře. Nejdříve bylo vytvořeno jeho grafické rozhraní (obr. 5), ve kterém si uživatel volí jednotlivé položky nastavení klasifikace a vizualizace.

Classificator			×
Layer	OLkraj	Show interval	CLASSIFICATOR
Field	hustota	Apply	HELP HOW TO USE:
Method	Natural Breaks (Jenks) Midle of Classification Mean	Close	 Select layer (only polygon) Select field to classify
Classes	6	•	3. Select method 4. If it is enabled. Select
Color Ramp	Black	▼ Takamuala	middle point of classification
Classification s Count: 397 Minimum: 4, Maximum: 9 Sum: 39971 Mean:100,6 Standard De	41 99,05 ,4599 838 eviation:110,4344	4,41 - 44,89 44,9 - 85,78 85,79 - 148,68 148,69 - 258,83 258,84 - 594,29 594,3 - 999,05	 5. Select number of classes 6. Select color ramp 7. Click on Show Interval to view breaks of intervals 8. Click on Apply to make selected intervals in map
			Contact: ivan.mail@post.cz

Obr. 5: Ukázka formuláře pro nastavení klasifikace

Uživatel má možnost si navolit pět parametrů. V nabídce Layer si vybere vrstvu, kterou využije pro klasifikaci. Skript využívá pouze polygonové vrstvy. Proto v případě, že není žádná polygonová vrstva načtená se v tomto poli neobjeví nic a skript nemůže pokračovat. Při změně položky Layer se automaticky aktualizuje nabídka Field. V této nabídce se navolí pole (atribut), které má být klasifikováno.

V nabídce Method se vybere požadovaná klasifikační metoda. Dostupné jsou metoda konstantních intervalů, geometrické klasifikace, Armstrong-Xiaova kritéria (Jenks Natural Breaks), kvantilů a mediánová kumulativní klasifikace.

V případě, že uživatel zvolí mediánovou kumulativní klasifikaci. Aktivuje se další nabídka výběru středu klasifikace (Middle of Classification), kde má na výběr z mediánu nebo průměru.

V dalším kroku si zvolí počet tříd. Jsou zde předdefinovány hodnoty od 1 do 12. Uživatel však podle potřeby může vepsat jakékoli číslo. V případě desetinného čísla ho skript automaticky zaokrouhlí na číslo celé.

Nakonec si uživatel navolí barevnou škálu, kterou použije. Na výběr je sedm barev: modrá (blue), černá (black), tyrkysová (cyan), zelená (green), červená (red) a žlutá (yellow).

Pod nabídkami vlevo je textové pole, kam se automaticky vypisují základní statistické charakteristiky vybraného atributu. Vpravo je druhé textové pole, kde se po stisku tlačítka Show interval zobrazí hodnoty intervalů podle navolených parametrů.

Dalším tlačítkem je tlačítko Apply. Toto tlačítko vyvolá klasifikaci a jednotlivé hodnoty přiřadí do odpovídajícího intervalu. Na základě příslušnosti do intervalu nastaví barvu polygonu v mapovém okně. Současně změní i legendu a přidá popis legendy v Table Of Contents, který je většinou umístěn na levé straně okna ArcMap 9.3.

Posledním tlačítkem je tlačítko Close, které provede uzavření celého formuláře. To lze samozřejmě provést i křížkem, který je typický pro aplikace Microsoft Windows, v pravé horní části formuláře.

V pravé části dialogového okna se nachází ještě jednoduchá nápověda, kde jsou popsány všechny kroky, které musí uživatel udělat, než začne klasifikovat data.

9.3. Tvorba skriptu

9.3.1. Definice proměnných

V prvním kroku se musely definovat všechny proměnné, které se ve skriptu používají. Tyto definice byly psány na začátek skriptu. To znamená před všechny funkce a události. První výraz Option Explicit vyžaduje, aby všechny proměnné byly deklarovány. Pokud tomu tak není, skript vypíše chybovou hlášku. Používá se především k zamezení chyb, kdy se špatně zapíše název proměnné.

```
Option Explicit
'variables declaration
Dim pMxDoc As IMxDocument
Dim pMap As IMap
Dim pEnumLayer As IEnumLayer
Dim pLayer As ILayer
Dim pFeatLayer As IGeoFeatureLayer 'IGeoFeatureLayer inherits
from IFeatureLayer. it mean it has same proterties and other
properties, methods
Dim pFc As IFeatureClass
Dim pCur As IFeatureCursor
Dim pFeat As IFeature
Dim pTable As ITable
Dim pFields As IFields
Dim pRow As IRow
'Help variable
Dim test As Variant
Dim layercount As Integer
Dim i As Integer
Dim LayerName As String
Dim FieldName As String
Dim ff As Variant
Dim rowcount As Integer
Dim max As Double
Dim min As Double
Dim sum As Double
Dim mean As Double
Dim StdDev As Double
Dim data()
Dim Inttxt()
Dim CountLayer As Double
Dim d1 As Double
Dim d2 As Double
Dim class As Integer
Dim midp As Double
```

```
'Variable for statistics
Dim pStatisticsResults As IStatisticsResults
'Variable for classifing and rendering
Dim pTableHistogram As ITableHistogram
Dim pHistogram As IHistogram
Dim DataValues As Variant
Dim DataFrequencies As Variant
Dim pClassify As IClassify
Dim Classes() As Double
Dim pClassBreaksRenderer As IClassBreaksRenderer
'Variable for create classificated layer
Dim pFillSymbol As IFillSymbol
Dim pRgbColor As IRgbColor
```

9.3.2. Inicializace formuláře

V inicializaci formuláře jsou vepsány všechny příkazy, které je nutné vykonat ihned po vyvolání formuláře. V tomto případě se jedná o přiřazení odkazu na objekty dokumentu a aktivní mapy do proměnných. Dále je zde naplnění comboboxů (označeny cb) hodnotami.

```
Private Sub UserForm_Initialize()
    'Setup references
   Set pMxDoc = Application.Document
   Set pMap = pMxDoc.FocusMap
    'Remove Items in cbLayer
    cbLayer.Clear
    'Add Items to cbLayer
        For i = 0 To pMap.layercount - 1
            Set pLayer = pMap.Layer(i)
            Set pFeatLayer = pLayer
            Set pFc = pFeatLayer.FeatureClass
            'Add only polygon layyers
            If pFc.ShapeType = esriGeometryPolygon Then
                LayerName = pMap.Layer(i).Name
                cbLayer.AddItem (LayerName)
            End If
       Next
    'Insert text to cbLayer
    cbLayer.Text = LayerName
    'Add Items to cbMethod
    cbMethod.AddItem ("Equal Interval")
    cbMethod.AddItem ("Geometrical Interval")
    cbMethod.AddItem ("Median's Cumulative")
    cbMethod.AddItem ("Natural Breaks (Jenks)")
   cbMethod.AddItem ("Quantile")
     'Insert text to cbClass
```

```
cbMethod.Text = "Median's Cumulative"
    'Add Items to cbDistr
    cbDistr.AddItem ("Mean")
    cbDistr.AddItem ("Median")
     'Insert text to cbDistr
    cbDistr.Text = "Mean"
    'Add colors as Items to cbColor
    cbColor.AddItem ("Blue")
    cbColor.AddItem ("Black")
    cbColor.AddItem ("Cyan")
    cbColor.AddItem ("Green")
    cbColor.AddItem ("Red")
    cbColor.AddItem ("Magenta")
    cbColor.AddItem ("Yellow")
    'Insert text to cbClass
    cbColor.Text = "Black"
End Sub
```

9.3.3. Procedura cbLayer_Change

V této proceduře jsou vepsány příkazy, které se provedou při změně hodnoty v comboboxu cbLayer (to znamená, když uživatele vybere jinou vrstvu pro klasifikaci). Tato procedura proběhne poprvé hned po inicializaci formuláře, protože už tam dojde ke změně (pokud je načtena alespoň jedna polygonová vrstva).

Hlavním cílem procedury je naplnění comboboxu cb_Atribute názvy všech polí (atributů) obsažených v tabulce vybrané vrstvy. Když dojde ke změně je předešlá nabídka vymazána a nahrazena novou.

```
Private Sub cbLayer_Change()
    'Remove Items in cbAtribute
   cbAtribute.Clear
    'Setup references
   Set pEnumLayer = pMap.Layers
   pEnumLayer.Reset
    'Find layer by name contain in cbLayer.Text
   Do
        Set pLayer = pEnumLayer.Next
   Loop Until pLayer.Name = cbLayer.Text
    'Insert this to the pFeatLayer
   Set pFeatLayer = pLayer
    'Setup references
   Set pFc = pFeatLayer.FeatureClass
   Set pFields = pFc.Fields
    'Add Items to cbAtribute from choose layer
   For i = 0 To pFields.FieldCount - 1
```

9.3.4. Procedura cbAtribute_Change

Procedura je velmi podobná principem předchozí. Reaguje však na změnu v comboboxu cbAtribute (uživatel změnil atribut pro klasifikaci). Poprvé proběhne na začátku po proběhnutí předešlé procedury. Potom jen když se změní hodnota v comboboxu cb_Atribute.

Procedura nastaví další proměnné a vypočítá základní statistické charakteristiky zvoleného atributu, které vypíše do textového pole txtStat.

```
Private Sub cbAtribute_Change()
    FieldName = cbAtribute.Text
    'Setup references
    Set pCur = pFc.Search(Nothing, False)
    Set pFeat = pCur.NextFeature
    'Number of row in table
   rowcount = pFeat.Table.rowcount(Nothing)
    Set pTable = pFeat.Table
    'Create histogram for selected Atribute
   Set pTableHistogram = New TableHistogram
   pTableHistogram.Field = FieldName
   Set pTableHistogram.Table = pTable
    'Count Statistics
   Set pStatisticsResults = pTableHistogram
   rowcount = pStatisticsResults.Count
   max = pStatisticsResults.Maximum
   min = pStatisticsResults.Minimum
   sum = pStatisticsResults.sum
   mean = pStatisticsResults.mean
    StdDev = pStatisticsResults.StandardDeviation
    'Insert text to txtStat
    txtStat.Text = "Count: " & Round(rowcount, 4) & vbNewLine &
"Minimum: " & Round(min, 4) & vbNewLine & "Maximum: " &
Round(max, 4) & vbNewLine & "Sum: " & Round(sum, 4) & vbNewLine
& "Mean:" & Round(mean, 4) & vbNewLine & "Standard Deviation:"
& Round(StdDev, 4)
End Sub
```

9.3.5. Procedura cbMethod_change

Tato procedura má za úkol při výběru mediánové kumulativní klasifiakce aktivovat výběr středu klasifikace. To znamená, od které hodnoty se budou intervaly počítat. Na výběr jsou možnosti: medián (median) nebo průměr (mean).

Druhou funkcí je, že při výběru zmíněné klasifikace v cbClass nastaví pouze sudé hodnoty počtu intervalů, protože tato metoda to vyžaduje.

```
Private Sub cbMethod_Change()
    cbClass.Clear
    If cbMethod.Text = "Median's Cumulative" Then
        'Add Items to cbClass (2-12 and only even number)
        For i = 2 To 12 Step 2
            cbClass.AddItem (i)
        Next
        'Make Option Button and its label enabled
        cbDistr.Enabled = True
        lblDistr.Enabled = True
    Else
        'Add Items to cbClass (1-12)
        For i = 1 To 12
            cbClass.AddItem (i)
        Next
        'Make Option Button and its label disabled
        cbDistr.Enabled = False
        lblDistr.Enabled = False
    End If
    'Insert text to cbClass
    cbClass.Text = "6"
End Sub
```

9.3.6. Procedury tlačítek

Po načtení předešlých kódů je možno začít klasifikovat data a vytvářet kartogramy. K tomu slouží již popsaná tlačítka vpravo.

Tlačítko Show Interval zobrazí intervaly v textovém poli. Procedura tlačítka (cmdInterval_Click) funguje tak, že nejdříve zavolá proceduru CreateClass (kapitola 9.3.6) a z ní získané intervaly zobrazí.

```
Private Sub cmdInterval_Click()
    'Call procedure CreateClass
    CreateClass
    'Create label of interval and it insert to txtInter
```

```
test = ""
   For i = 0 To pClassBreaksRenderer.BreakCount - 1
       pClassBreaksRenderer.Break(i) = Classes(i + 1)
        If i = 0 Then
            pClassBreaksRenderer.Label(i) =
CStr(Round(Classes(i), 2)) & " - " & CStr(Round(Classes(i + 1),
2))
        Else
            pClassBreaksRenderer.Label(i) =
CStr(Round(Classes(i), 2) + 0.01) & " - " &
CStr(Round(Classes(i + 1), 2))
        End If
        test = test + vbNewLine + pClassBreaksRenderer.Label(i)
   Next
     txtInter.Text = test
End Sub
```

Dalším tlačítkem je Apply. To na základě klasifikace přiřadí danou barvu polygonu v mapě a vytvoří legendu. Stisk tlačítka vyvolá proceduru cmdApply_Click, která volá dvě procedury CreateClass (kapitola 9.3.6) a CreateMap (kapitola 9.3.7).

```
Private Sub cmdApply_Click()
    'Call procedures CreateClass and CreateMap
    CreateClass
    CreateMap
End Sub
```

Posledním tlačítkem je tlačítko Close, které vyvolá proceduru cmdClose_Click a zavře formulář.

```
Private Sub cmdClose_Click()
End
End Sub
```

9.3.7. Procedura CreateClass

Tato procedura by se dala nazvat mozkem celého skriptu. Na základě volby klasifikační metody vypočítá hranice jednotlivých intervalů a následně do nich rozloží data. Tato procedura je volána jak stiskem tlačítka Show interval tak i tlačítkem Apply.

Funguje tak, že nejdříve vytvoří z vybraných dat histogram. Potom určí vybranou metodu a aplikuje ji na data.

```
Private Sub CreateClass()
   Set pTableHistogram = New TableHistogram
   pTableHistogram.Field = FieldName
   Set pTableHistogram.Table = pTable
    ' Get values and frequencies from the histogram
   Set pHistogram = pTableHistogram
   pHistogram.GetHistogram DataValues, DataFrequencies
    ' Selection of method from cbMethod
   If cbMethod.Text = "Natural Breaks (Jenks)" Then
        Set pClassify = New NaturalBreaks
   ElseIf cbMethod.Text = "Equal Interval" Then
        Set pClassify = New EqualInterval
    ElseIf cbMethod.Text = "Geometrical Interval" Then
        Set pClassify = New GeometricalInterval
    ElseIf cbMethod.Text = "Quantile" Then
        Set pClassify = New Quantile
    ElseIf cbMethod.Text = "Geometrical Interval" Then
        Set pClassify = New GeometricalInterval
    ElseIf cbMethod.Text = "Median's Cumulative" Then
        Set pClassify = New GeometricalInterval
   End If
   pClassify.SetHistogramData DataValues, DataFrequencies
   pClassify.Classify Val(cbClass.Text)
   If cbMethod.Text = "Median's Cumulative" Then
        MedianClass
   Else
        Classes = pClassify.ClassBreaks
   End If
    ' Create BreaksRenderer --> Get Break Value
   Set pClassBreaksRenderer = New ClassBreaksRenderer
   With pClassBreaksRenderer
        .Field = FieldName
        .BreakCount = Val(cbClass.Text)
        .MinimumBreak = Classes(0)
   End With
End Sub
```

9.3.8. Procedura CreateMap

Procedura je volána při stisku tlačítka Apply. Vychází z klasifikačních tříd vytvořených předešlou procedurou. Na základě vybrané barevné škály přiřadí barvu každé hodnotě (resp. polygonu). Potom vytvoří novou legendu z použitých barev a jako popis vloží hranice intervalů. Nakonec provede obnovení mapového pole a legendy v ArcMap.

```
Private Sub CreateMap()
   For i = 0 To pClassBreaksRenderer.BreakCount - 1
        ' Setup symbol
        Set pFillSymbol = New SimpleFillSymbol
```

```
Set pRgbColor = New RgbColor
    ' Choose and create color ramp
        If cbColor.Text = "Red" Then
            pRgbColor.Red = (((205 /
pClassBreaksRenderer.BreakCount) * i) + 50)
            pRgbColor.Green = 0
            pRgbColor.Blue = 0
        ElseIf cbColor.Text = "Green" Then
            pRgbColor.Red = 0
            pRgbColor.Green = (((205 /
pClassBreaksRenderer.BreakCount) * i) + 50)
            pRqbColor.Blue = 0
        ElseIf cbColor.Text = "Blue" Then
            pRqbColor.Red = 0
            pRgbColor.Green = 0
            pRgbColor.Blue = (((205 /
pClassBreaksRenderer.BreakCount) * i) + 50)
        ElseIf cbColor.Text = "Black" Then
            pRqbColor.Red = (((205 /
pClassBreaksRenderer.BreakCount) * i) + 50)
            pRqbColor.Green = (((205 /
pClassBreaksRenderer.BreakCount) * i) + 50)
            pRgbColor.Blue = (((205 /
pClassBreaksRenderer.BreakCount) * i) + 50)
        ElseIf cbColor.Text = "Yellow" Then
            pRgbColor.Red = (((205 /
pClassBreaksRenderer.BreakCount) * i) + 50)
            pRgbColor.Green = (((205 /
pClassBreaksRenderer.BreakCount) * i) + 50)
            pRgbColor.Blue = 0
        ElseIf cbColor.Text = "Magenta" Then
            pRgbColor.Red = (((205 /
pClassBreaksRenderer.BreakCount) * i) + 50)
            pRgbColor.Green = 0
            pRgbColor.Blue = (((205 /
pClassBreaksRenderer.BreakCount) * i) + 50)
        ElseIf cbColor.Text = "Cyan" Then
            pRqbColor.Red = 0
            pRgbColor.Green = (((205 /
pClassBreaksRenderer.BreakCount) * i) + 50)
            pRqbColor.Blue = (((205 /
pClassBreaksRenderer.BreakCount) * i) + 50)
        Else
            pRgbColor.Red = (((205 /
pClassBreaksRenderer.BreakCount) * i) + 50)
            pRgbColor.Green = 0
            pRqbColor.Blue = 0
        End If
        ' Define color of Fill Symbol (color is insert reverse
to low intervals have lighter color)
        pFillSymbol.Color = pRgbColor
     pClassBreaksRenderer.Symbol(pClassBreaksRenderer.BreakCou
nt - 1 - i) = pFillSymbol
        pClassBreaksRenderer.Break(i) = Classes(i + 1)
```

```
'Create label of intervals
        If i = 0 Then
            pClassBreaksRenderer.Label(i) =
CStr(Round(Classes(i), 2)) & " - " & CStr(Round(Classes(i + 1),
2))
        Else
            pClassBreaksRenderer.Label(i) =
CStr(Round(Classes(i), 2) + 0.01) & " - " &
CStr(Round(Classes(i + 1), 2))
        End If
   Next
    'Change symbology of layer (therefore we use
IGeoFeatureLayer) and refresh map document
    Set pFeatLayer.Renderer = pClassBreaksRenderer
    pMxDoc.ActiveView.PartialRefresh esriViewGeography,
pFeatLayer, Nothing
    pMxDoc.UpdateContents
End Sub
```

9.3.9. Procedura MedianClass

Tato procedura je volána pouze, když je vybrána klasifikace pomocí mediánové kumulativní metody. Prvním krokem je načtení dat z vybraného pole tabulky.

Dále podle zvoleného středu klasifikace (průměr, medián) vypočítá tuto hodnotu z načtených dat.

Při výpočtu mediánu je potřeba setřídit hodnoty podle velikosti a určit hodnotu, která je uprostřed. Tato hodnota je medián.

Pro seřazení byl použit princip SnakeSort, který je podle [27] hodnocen jako nejrychlejší. Zdrojový kód SnakeSort kvůli jeho délce zde není uveden, ale je dostupný na http://www.vbforums.com/showthread.php?t=473677.

Po výpočtu středu dopočítá všechny hranice intervalů které uloží do proměnné Classes, s kterou dále pracuje procedura CreateClass (kapitola 9.3.7).

```
Private Sub MedianClass()
   'if median is selected do midp = median
   If cbDistr.Text = "Median" Then
        ff = pTable.FindField(FieldName)
        'array range
        ReDim data(rowcount)
        'All values load to array as Double (CDbl)
```

```
For i = 0 To rowcount - 1
            data(i) = CDbl(pTable.GetRow(i).Value(ff))
            Set pRow = pTable.GetRow(i)
            data(i) = CDbl(pRow.Value(ff))
       Next i
        'call function SnakeSort1 to sort data
       SnakeSort1 (data)
    'if mean is selected do midp = mean
   Else
       midp = mean
   End If
    'define array size
   ReDim Classes(Val(cbClass.Text) + 1)
    'count breaks of intervals
   class = Val(cbClass.Text) / 2
   d1 = (2 * (midp - min)) / (class * (class + 1))
   d2 = (2 * (max - midp)) / (class * (class + 1))
    'take values of breaks of intervals to array Classes
   Classes(0) = min
   For i = 1 To class + 1
        Classes(i) = Classes(i - 1) + ((class - (i - 1)) * d1)
   Next
   For i = i - 1 To Val(cbClass.Text) - 1
        Classes(i) = Classes(i - 1) + ((i - class) * d2)
   Next
   Classes(i) = max
End Sub
```

9.4. Vhodnost použití metody

Metoda mediánové kumulativní klasifikace byla, jako ostatní metody, vyzkoušena na testovacích datech. Metoda vykazovala celkem dobré výsledky u obou datasetů.

Z testování je zřejmé, že optimální data pro tuto metodu by neměla obsahovat osamocené extrémní hodnoty. Tyto hodnoty zapříčiňují velkou šířku intervalu a tím pádem velké množství dat v prostředních intervalech a minimum dat v intervalech krajních. To je vidět u dat s normálním rozdělením (příloha 11) u prvního intervalu, který obsahuje 3 obce. Logaritmické rozdělení (příloha 10) má v šestém intervalu 4 obce. Naopak čtvrtý interval obsahuje velké množství obcí.

I když jsou zde tyto nedostatky, vždy je splněna podmínka, že v první polovině intervalů je stejný počet hodnot jako v druhé.

10. Diskuse

V této práci je výčet klasifikačních metod kvantitativních dat v geografických informačních systémech. Jejich dostupnost byla testována v šesti softwarech. V budoucnu by bylo možné takto otestovat více softwarů a vytvořit i složitější hodnotící metody jako příklad by mohla sloužit kniha [3], kde je popsáno více metod i s ukázkami.

Vhodnost metod byla hodnocena z hlediska statistického. Tím se myslí, že byla testována dvě rozdělení dat (normální a logaritmické) a k nim vizuálně byly určeny vhodné klasifikační metody.

Jak už bylo zmíněno nejedná se zrovna o objektivní metodu. Bylo by dobré vymyslet matematicky podloženou metodu, která by testovala vhodné rozdělení do jednotlivých klasifikačních tříd. Tím by se odstranila subjektivita hodnotitele.

Vhodnost metod by v budoucnu mohla být rozšířena o další pohledy než podle typu použitých dat. Tato metoda byla užita proto, že podle [24] v dnešní době pracují s geografickými informačními systémy a tudíž i s geografickými daty nekartografové. Odborníci v jiných disciplínách, kteří potřebují vizualizovat své výsledky. Tito lidé neznají principy tvorby kartografických děl.

Většinou si ani neuvědomí, že používají data, která nejsou v normálním rozdělení. Tyto data však potřebují také prezentovat v grafické podobě. Většinou tito lidé nemají chuť, a ani čas tvořit vlastní stupnici. Tato práce je proto chce upozornit na to, že existuje více klasifikačních metod než pouze předvolená. Dále jim nabídla řešení, jaké metody se dají použít u jednotlivých typů rozdělení. Je samozřejmostí, že u jiných dat nemusí být vždy všechny metody vhodné, ale "příležitostným kartografům" pomohou s výběrem.

Hodnotit rozdělení a následný kartografický způsob lze z pohledu mnoha oborů. Zde se vybral pohled statistický, resp. matematický. Statistik bude mít jiné požadavky než třeba demograf, urbanista atd.

Toto by mohlo být téma další diplomové práce. Kdy by se metody hodnotily podle účelu mapy. Účel se určuje podle toho, komu mapa bude sloužit a k čemu ji bude uživatel potřebovat. [24]

Postup by mohl vypadat následovně. Rozdaly by se mapové výstupy s různými klasifikacemi zástupcům jednotlivých profesí, třeba i veřejnosti, a ti by ze svého pohledu určili vhodné metody a důvod proč jsou vhodné. Důvod by vyjádřili tím, co se z dané mapy dozvěděli. A tato informace by pro ně byla dostatečná.

Například demograf bude potřebovat z mapy vytyčit regiony s malou, střední a velkou hustotou obyvatel (na příkladě testovaných dat by pro něho byla nejvhodnější metoda kvantilů). Urbanistu, který má plánovat výstavbu, budou zajímat pouze regiony s největší hustotou, aby mohl v těchto místech navrhovat rozvoj měst. Pro něho by byla nejvhodnější metoda geometrické klasifikace nebo Armstrong-Xiaovo kritérium.

Takto by se dalo určit, co která profese od mapových výstupů očekává a jaké jsou pro ni nejvhodnější výsledky.

Místo účelu by se mohla hodnotit vhodnost podle funkce. Např. funkce informační, orientační, klasifikační, rozhodovací [24].

Jako poslední cíl práce byla implementace mediánové kumulativní klasifikace do ArcGIS 9.3 skriptem psaným v jazyce MS Visual Basic 6.3. Součástí skriptu jsou i další metody, které jsou dostupné v programu ArcGIS 9.3.

Zdrojový kód je volně přístupný, proto se do něho dají vepsat další metody a tím ho rozšířit. Třeba zmíněnou metodou Caspall-Jenksova kritéria a dalšími.

11. <u>Závěr</u>

Tato diplomová práce shrnuje a popisuje v současné době dostupné klasifikační metody kvantitativních dat v geografických informačních systémech.

Dostupnost metod byla testována v šesti GIS programech. Tři byly komerční (ArcGIS 9.3, Geomeedia Professional 6.0 a AutoCAD Map 3D 2009) a zbylé programy byly open source (GRASS 6.4.0, Qunatum GIS 1.0.1 a uDig 1.1.0-SC3).

Všechny tyto programy byly porovnávány s ohledem na klasifikační metody. Hlavním kritériem byl počet implementovaných metod. Software ArcGIS 9.3 umí pracovat s nejvíce metodami, a to sedmi. Na druhém místě se umístil AutoCAD Map 3D a GRASS s pěti metodami. GRASS je tedy nejlepší open source software pro klasifikaci dat z testovaných programů.

Vhodnost užití metod byla testována pro dvě sady dat s logaritmickým a normálním rozdělením. Metodou bylo vizuální hodnocení klasifikačních metod. To nezaručuje objektivní výsledky, ale je dostatečné pro určení vhodných metod.

Jako nejvhodnější klasifikační metoda pro logaritmické rozdělení byla určena metoda kvantilů. Dalšími vhodnými metodami byly geometrická klasifikace a Armstrong-Xiaovo kritérium (Natural Breaks Jenks).

Pro vybraná data v normálním rozdělení byla nejlepší metoda kvantilů. Dobré výsledky také opět poskytovaly metody geometrické klasifikace a Armstrong-Xiaovo kritérium

Metoda konstantních intervalů byla pro oba druhy dat nevhodná. Tato metoda by byla vhodná pro data v konstantním rozložení.

Dále byla implementována nová klasifikační metoda do ArcGIS 9.3. Byl napsán skript v jazyce MS Visual Basic 6.3. Skript nabízí kromě metod v ArcGIS již začleněných i novou metodu, mediánovou kumulativní klasifikaci. Ta funguje tak, že tvoří stejný počet tříd na obě strany od mediánu nebo průměru (podle nastavení). Šířka třídy se směrem od mediánu (průměru) zvětšuje.

Skript s novou klasifikační metodou byl vložen do toolbaru. Celý skript je okomentován v anglickém jazyce. Toolbar bude možné stáhnout na webových stránkách magisterské práce a na CD-ROMu na Katedře geoinformatiky Univerzity Palackého v Olomouci.

12. Seznam použitých zdrojů

Literatura:

- BAYER, T. (2008): Algoritmy v digitální kartografii. Praha: Univerzita Karlova v Praze, 251 s.
- [2] BLÁHA, D.,J. (2009): Various ways of assessment of cartographic works, ICA symposium on Cartography for Central and Eastern Europe, Vienna, 11 s.
- [3] DOBEŠOVÁ, Z. (2009): Hodnocení kartografické funkcionality geografických informačních systémů. Univerzita Palackého v Olomouci, Olomouc, 131 s.
- [4] DOBEŠOVÁ, Z. (2007): AutoCAD Map 3D Praktická uživatelská příručka.
 Computer Press, a. s., Brno, 238 s.
- [5] KAŇOK, J. (1999): *Tematická kartografie*. Ostravská univerzita, Ostrava, 318 s.
- [6] KRAAK, M., J., ORMELING, F (2003).: Cartography, Visualization of Geospatial data, Second Edition. Prentice Hall, London, 205 s.
- [7] MERTENOVÁ, M. (2007): Možnosti tvorby kartografických výstupů v prostředí Open Source GIS nástrojů. bakalářská práce, Vysoká škola báňská – Technická univerzita Ostrava, Hornicko-geologická fakulta, Institut geoinformatiky, Ostrava, 70 s.
- [8] MORKES, D. (2000): Visual Basic 6.0. Computer Press, Praha, 165 s.

- [9] PRAVDA, J., KUSENDOVÁ, D. (2004): Počítačová tvorba tematických map. Bratislava, Univerzita Komenského v Bratislavě ve vydavatelství UK, 264 s.
- [10] SLOCUM, T., A., McMASTER, R., B., KESSLER, F., C., HOWARD, H.,
 H. (2004): *Thematic Cartography and geographic Visualization*, Prentice Hall, 518 s.
- [11] TUFTE, E., R. (2007): *The Visual Display of Quantitative Information*. Graphic Press, Cheshire, 197 s.
- [12] TUFTE, E., R. (2005): Visual Explanations. Graphic Press, Cheshire, 156 s.
- [13] VEVERKA, B. (2004): Topografická a tematická kartografie 10. Praha, Vydavatelství ČVUT, 220 s.
- [14] VOŽENÍLEK, V. (1999): Aplikovaná kartografie I. tématické mapy.
 Olomouc, UP, 170 s.
- [15] VOŽENÍLEK, V. (2000): GIS I pojetí, historie, základní komponenty.
 Olomouc, UP, 173 s.
- [16] VOŽENÍLEK, V. (2002): Diplomové práce z geoinformatiky. Olomouc,
 Vydavatelství Univerzity Palackého, 31 s.

Elektronické zdroje:

[17] *About QGIS* [online]. c2009 [cit. 2010-04-12]. Dostupný z WWW: http://qgis.org/en/about-qgis.html.

- [18] ArcGIS Desktop ARCDATA PRAHA [online]. c2010 [cit. 2010-04-12]. Dostupný z WWW: http://www.arcdata.cz/produkty-a-sluzby/software/esri/arcgis-desktop/>.
- [19] ČERBA, O. *Kartogramy.pdf* [online]. 2009 [cit. 2010-04-12]. Dostupné z
 WWW: http://www.gis.zcu.cz/studium/tka/Slides/kartogramy.pdf>.
- [20] *ESRI Support* [online]. c2010 [cit. 2010-04-12]. Dostupný z WWW: .
- [21] GNU General Public License In Wikipedia : the free encyclopedia [online].
 St. Petersburg (Florida) : Wikipedia Foundation, 20.12.2007, 5.4.2010 [cit. 2010-04-12]. Dostupné z WWW: http://cs.wikipedia.org/wiki/GNU_General_Public_License>.
- [22] GRASS GIS: about [online]. c1999-2007 [cit. 2010-04-12].Dostupný z WWW: http://grass.itc.it/intro/general.php.
- [23] Intergraph / Hlavní stránka [online]. c2010 [cit. 2010-04-12].
 Dostupný z WWW: ">http://www.intergraph.com/global/cz/.
- [24] KAŇOK, J. (2008): Chyby v mapových výstupech ovlivňující čtení, interpretaci znázorněných jevů a následně rozhodovací proces. In GIS Ostrava 2008. Ostrava, [cit. 2010-04-17]. Dostupné z WWW: <http://gis.vsb.cz/GIS_Ostrava/GIS_Ova_2008/sbornik/Lists/Papers/050.pdf >.
- [25] KRAAK, M., J. (2002): Current trends in visualisation of geospatial data with special reference to cartography. In *Indian Cartographer*. [cit. 2009-11-05]. Dostupné z WWW: http://www.nrac.wvu.edu/classes/RESM440

/MaterialsByWeek/Week4/Kraak_2002_CurrentTrendsinVisualization-Cartography.pdf>.

- [26] Udig : Home [online]. c2009 [cit. 2010-04-12]. Dostupný z WWW: http://udig.refractions.net/>.
- [27] VB6: Sorting algorithms (sort array, sorting arrays) VBForums [online].
 c2000 2010 [cit. 2010-03-18]. Dostupný z WWW: http://www.vbforums .com/showthread.php?t=473677>

<u>Summary</u>

This Diploma Theses summarizes and describes the classification methods of the quantitative data in the geographical information systems that are available in present days.

The accessibility of the methods was proved in six GIS programs. Three of them were commercial (ArcGIS 9.3, Geomedia Professional 6.0 and AutoCAD Map 3D 2009) and the rest of them was open sources (GRASS 6.4.0, Quantum GIS 1.0.1 and uDig 1.1.0-SC3).

All of these programs were compared with regard to the classificaton methods. The main criterion was the number of the implemented methods. Software ArcGIS 9.3 can work with the highest number of methods – with seven methods. The second place occupies AutoCAD Map 3D and GRASS that can work with five methods. It means that GRASS is the best open source software for the data classification of all tested programs.

The suitability of the methods use was tested for two sets of data with logarithmic and normal distribution. The method was visual evaluation of classifiation methods. This doesn't warrant the objective results, but it is sufficiently for the suitable methods setting.

As the suitable classification methods for logaritmic division were set: the quantile method.

The quantile method and the Armstrong-Xiaovo criterion (Natural Breaks Jenks) were the best for the chosen data in the normal division. The geometrical classification method provided the satisfactory results. The results show that it is better to use more than one suitable method for the data in normal division.

The method of the equal intervals was unsuitable for both types of the data. This method would be suitable for the data in the constant distribution. The new classification method was implemented to the ArcGIS 9.3. The script was written in the MS Visual Basic 6.3. The script offers not only the Median Cumulative Classification, but also the methods which are already integrated.

The script with the new classification method was inserted to the toolbar. The whole script is commented in english. The toolbar and the source code will be able to download on the website of the diploma thesis and on the CD-ROM in Department of geoinformatics in Palacký University in Olomouc.

<u>PŘÍLOHY</u>

<u>Seznam příloh</u>

- Příloha 1: Realizace metod klasifikace kvantitativních dat v GIS (CD-ROM)
- Příloha 2: Hustota zalidnění. Klasifikační metoda konstantní intervaly při logaritmickém rozdělení.
- Příloha 3: Hustota zalidnění. Klasifikační metoda kvantily při logaritmickém rozdělení.
- Příloha 4: Hustota zalidnění. Klasifikační metoda geometrická klasifikace při logaritmickém rozdělení.
- Příloha 5: Hustota zalidnění. Klasifikační metoda Armstrong-Xiaovo kritérium při logaritmickém rozdělení.
- Příloha 6: Podíl žen. Klasifikační metoda konstantní intervaly při normálním rozdělení.
- Příloha 7: Podíl žen. Klasifikační metoda kvantily při normálním rozdělení.
- Příloha 8: Podíl žen. Klasifikační metoda geometrická klasifikace při normálním rozdělení.
- Příloha 9: Podíl žen. Klasifikační metoda Armstrong-Xiaovo kritérium při normálním rozdělení.
- Příloha 10: Hustota zalidnění. Klasifikační metoda mediánové kumulativní klasifikace při logaritmickém rozdělení.
- Příloha 11:Podíl žen. Klasifikační metoda mediánové kumulativní klasifikace při normálním rozdělení.