



# **Geoinformatika**

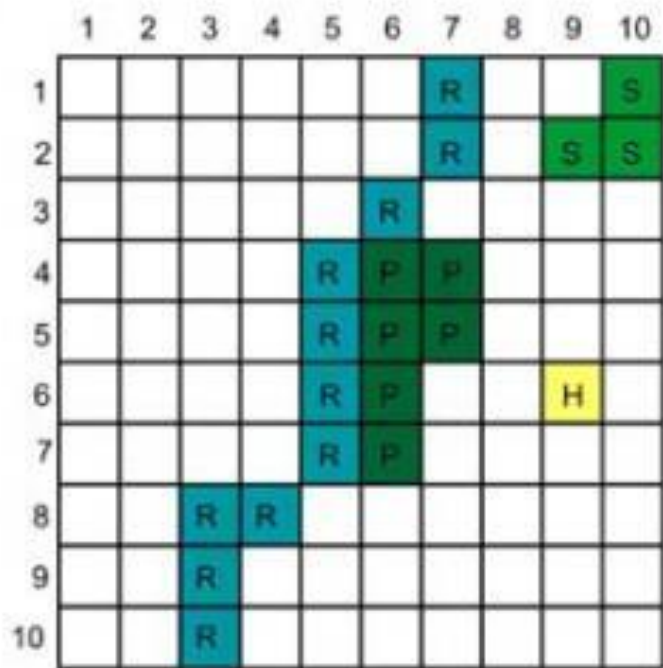
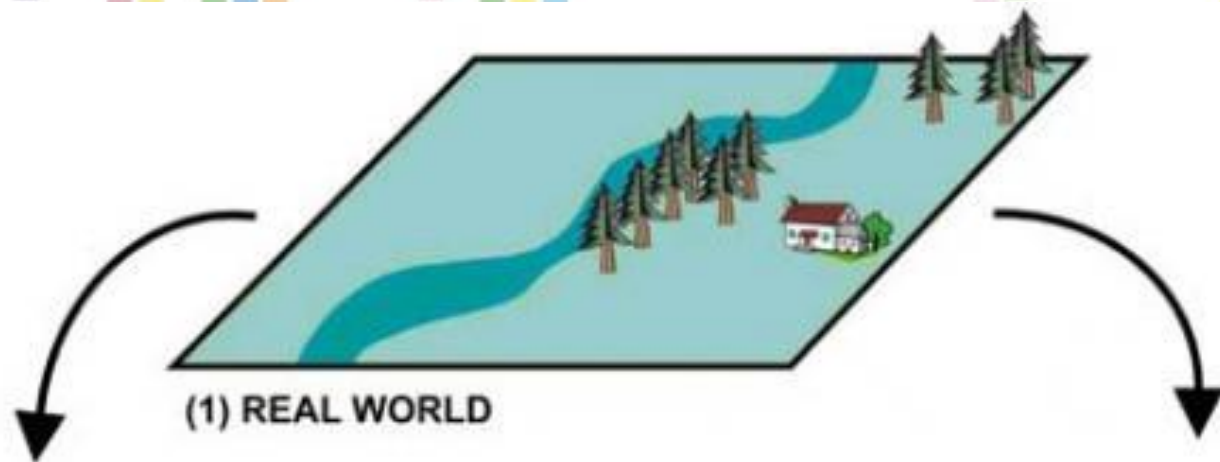
## **III. Rastrový datový model**

**jaro 2020**

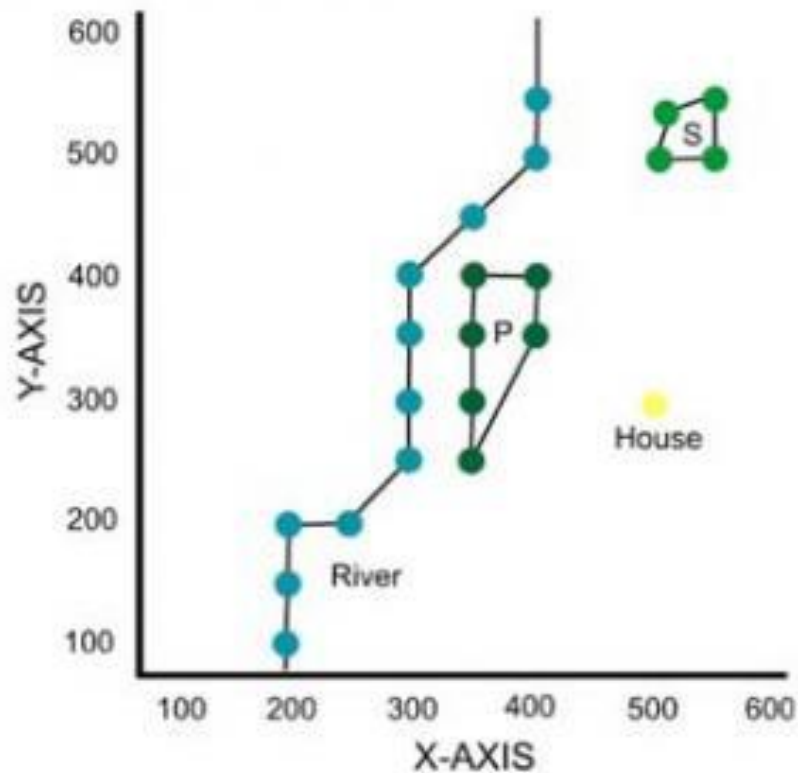
**Petr Kubíček**

**kubicek@geogr.muni.cz**

**Laboratory on Geoinformatics and Cartography (LGC)  
Institute of Geography  
Masaryk University  
Czech Republic**



(2) RASTER REPRESENTATION



(3) VECTOR REPRESENTATION

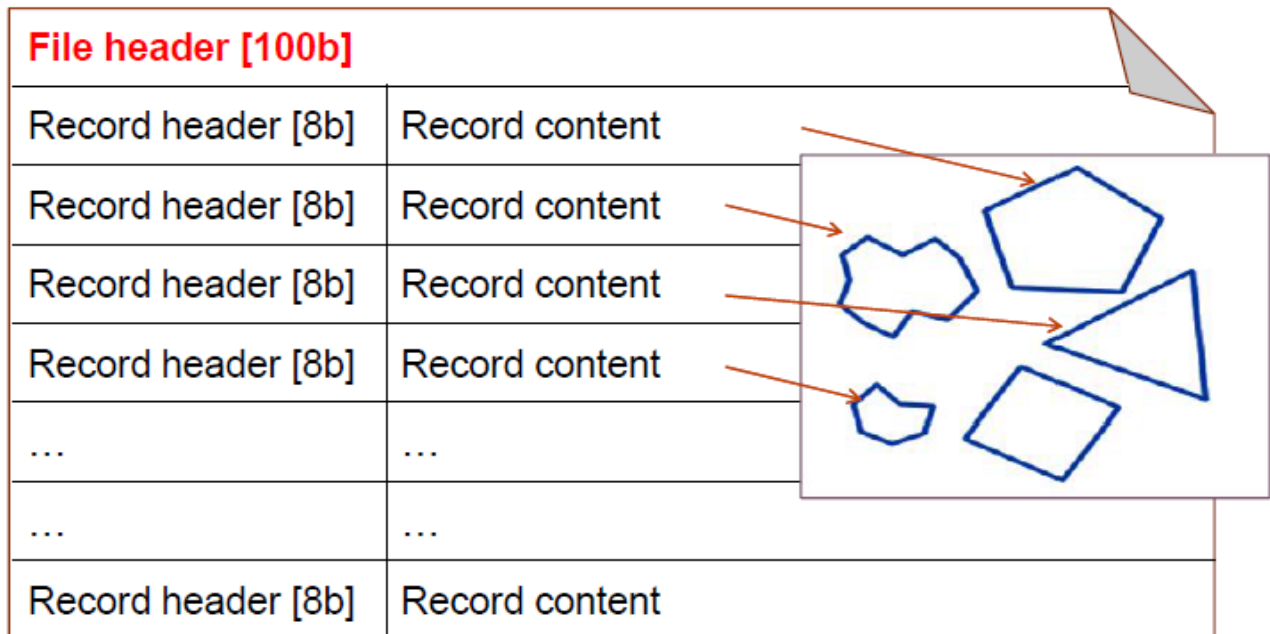
# ArcGIS Shapefile

- Jeden soubor obvykle reprezentuje jeden typ mapového prvku, např. silnice, jezera, obce.
- Shapefile specifikuje i další pomocné soubory.
- „**Jméno.přípona**“ prefix zůstává stejný, přípona se mění:
- **Povinné**
  - **.shp** – samotný hlavní soubor s geodaty (geometrie).
  - **.shx** – indexový soubor (posun vůči počátku souboru, délka záznamu).
  - **.dbf** – soubor s atributy resp. popisné data.
- **Nepovinné**
  - **.prj** – zdrojový souřadnicový systém.
  - WKT-string  
GEOGCS["GCS\_WGS\_1984",DATUM["D\_WGS\_1984",SPHEROID["WGS\_1984",6378137,298.257223563]],PRIMEM["Greenwich",0],UNIT["Degree",0.017453292519943295]]
  - **.cpg** – specifikuje kódování v dbf souboru.
  - UTF-8.
  - **.mxd – map exchange format** – ukládá všechny související informace – mapa, symbologie, hyperlinky, mapové okno...

# Struktura \*.SHP

- Geometrický prvek v záznamu - **shape**
- Samotnou geometrii shape ukládá jako **sekvenci bodů** (např. GPS souřadnic).
- Nedefinuje topologickou strukturu.
- Jeden záznam shape – jeden řádek v attributech.

Point

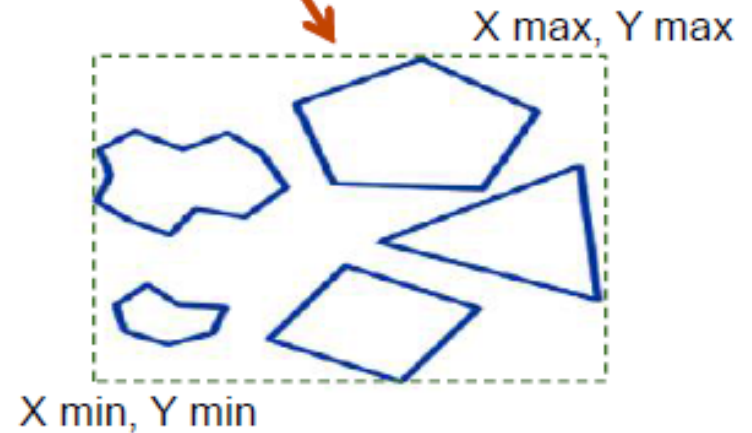




# Hlavička souboru \*.SHP

0	4	24	28	32	36	68	84
int	int	int	int	int	4 x double	2x double	2 x double
File code vždy 94440	nepouž.	délka souboru	verze	Typ shape	MBR X min Y min X max Y max	Z min Z max	M min M max

0	NullShape	15	PolygonZ
1	Point	18	MultiPointZ
3	PolyLine	21	PointM
5	Polygon	23	PolyLineM
8	MultiPoint	25	PolygonM
11	PointZ	28	MultiPointM
13	PolyLineZ	31	MultiPatch





# Hlavička/obsah záznamu

File header [100b]

<b>Record header [8b]</b>	Record content
Record header [8b]	Record content

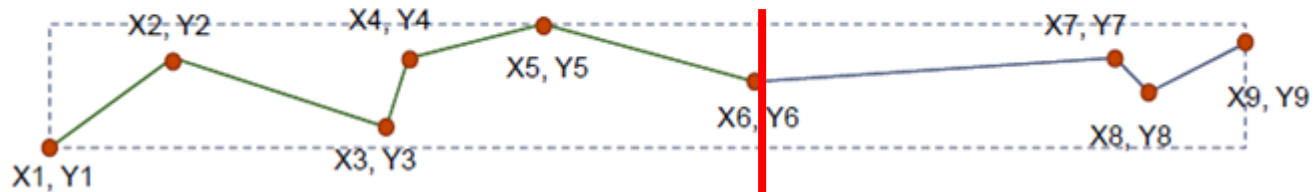
<b>0</b>	<b>4</b>
int	int
číslo záznamu	délka záznamu

File header [100b]

Record header [8b] **Record content**

<b>0</b>	<b>4</b>	<b>36</b>	<b>40</b>	<b>44</b>	<b>X</b>	<b>EOF</b>
int	4 x double	int	int	Int[]	Point[] (2 x double na 1 bod)	
<b>Shape type</b>	<b>MBR</b>	<b>Počet částí</b>	<b>Počet bodů</b>	<b>Indexy na části</b>	<b>Body</b>	$X = 44 + 4 * \text{počet částí}$

3	Xmin, Ymin, Xmax, Ymax	2	9	[0,5]	[X1,Y1], [X2,Y2], ..., [X6,Y6],..., [X9,Y9]
---	------------------------	---	---	-------	---





# Atributová data \*.dbf

- Standartní DBF soubor (tabulka)
- Ke každému záznamu existuje právě jeden řádek v tabulce ve stejném pořadí jako ve zdrojovém shapefile.
- Stejný prefix jako zdrojový shapefile.
- Kódování uloženo v **.cfg** souboru.
- Velké množství dat, redundance.

	SHAPE_ID	LINK_ID	ST_NAME	FEAT_ID	ST_LANGCD	NUM_STNM	ST_NM_F
▶	0	565809744	CINGROVA	1248477527	CZE	2	
	1	565809752		0		0	
	2	565809753	CINGROVA	1248477527	CZE	2	
	3	565809754		0		0	
	4	565809755	CINGROVA	1248477527	CZE	2	
	5	565809756	CINGROVA	1248477527	CZE	2	
	6	565809757	56	1410701498	CZE	2	
	7	565809758	56	1410701498	CZE	2	

# Shapefile - shrnutí

## + **výhody**

- Neukládá topologii dat
- Snadná editace bodů
- Rychlá vizualizace geodat
- Jednoduše pochopitelná struktura
- Podpora v GIS softwarech
- Snadná projekce do jiných souřadnicových systémů

## - **nevýhody**

- Neukládá topologii dat
- Redundance dat (např. body sousedících polygonů)
- Manipulace s detailní shapefile (až 100MB soubor) je pomalá.
- Špatná podpora Unicode.



# Vektorová data

## Výhody

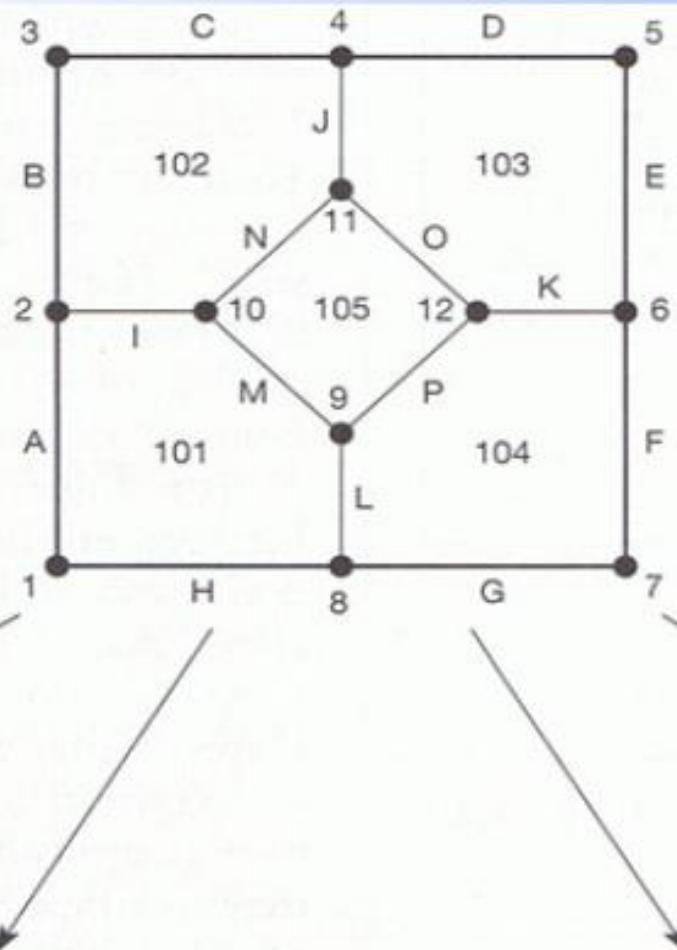
- lze pracovat s jednotlivými objekty jako se samostatnými celky;
- menší náročnost na paměť;
- dobrá reprezentace jevové struktury dat;
- **vysoká geometrická přesnost**
- kvalitní grafika, přesné kreslení, znázornění blízké mapám;
- jednoduché vyhledávání, úpravy a generalizace objektů a jejich atributů.

## Nevýhody

- výpočtová náročnost (problémy při náročných analytických operacích);
- komplikovanost datové struktury;
- složitější odpovědi na polohové dotazy;
- obtížná tvorba překryvů vektorových vrstev (overlay)
- problémy při modelování a simulaci jevů.



# Kahoot 😊



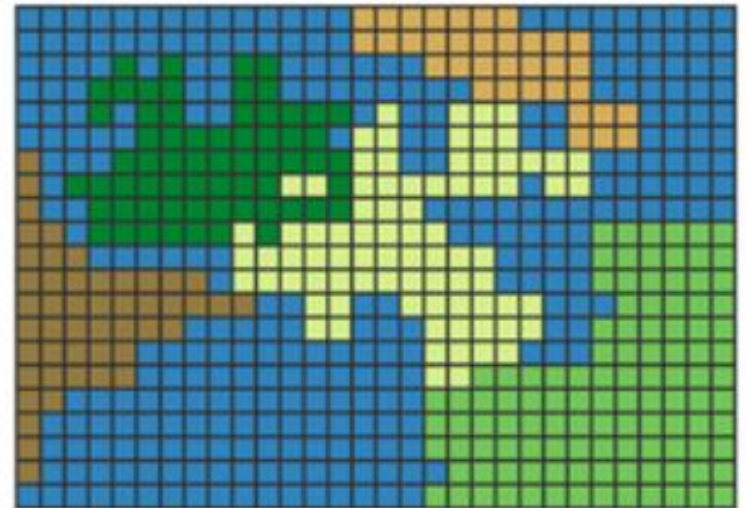
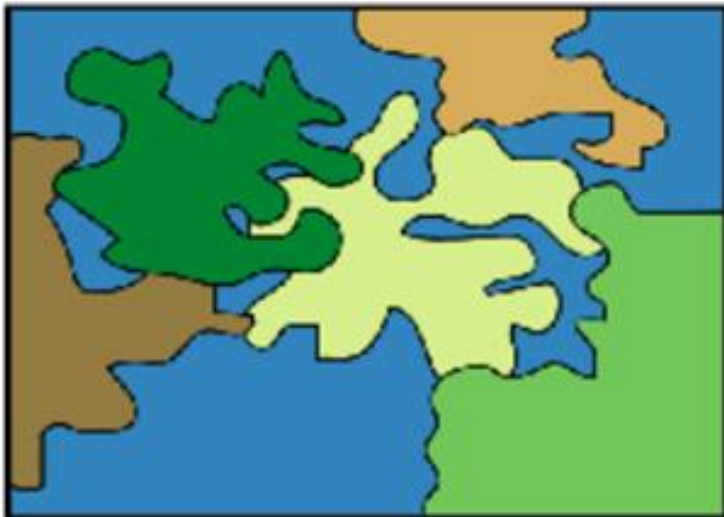
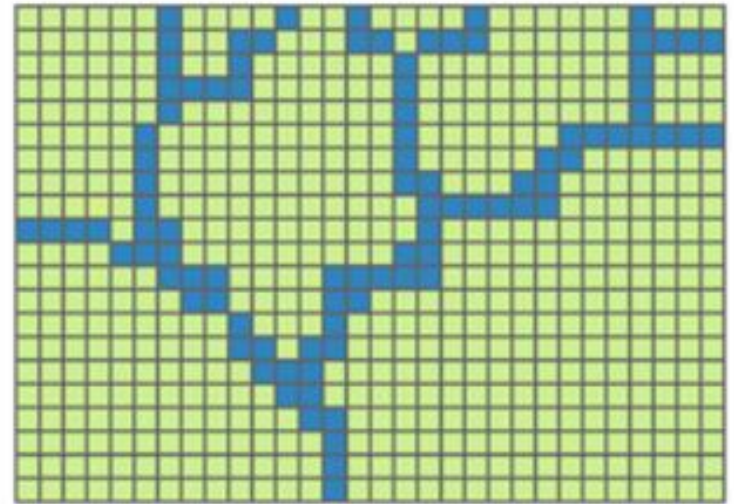
Node file		
ID	X	Y
1	0	0
2	0	10
3	0	20
⋮	⋮	⋮
12	14	10

Polygon structure file	
ID	Chain/Segment list
101	A, I, M, L, H
102	B, C, J, N, I
103	D, E, K, O, J
104	F, G, L, P, K
105	M, N, O, P

Chain/Segment file					
ID	Start-node	End-node	Left-poly	Right-poly	Length
A	1	2	Outside	101	10
B	2	3	Outside	102	10
C	3	4	Outside	102	10
⋮	⋮	⋮	⋮	⋮	⋮
P	12	9	104	105	4

Polygon attribute file		
ID	VAR 1 (Name)	VAR 2 (Area)
101	Cars	96
102	Cars	96
103	Staff	96
104	Buses	96
105	Info kiosk	16

# Vektor vs. rastr



# Rastrová reprezentace

- **Zaměřuje se na lokalitu jako na celek**
- **Používá se pro reprezentaci jevů, které plošně pokrývají celou oblast, případně se i spojitě mění.**
- **Používá se i pro rasterizované vektorové vrstvy, pokud je následná analýza jednodušší nad rastrem.**
- **RAVE - VERA**

# Rastrová reprezentace

- Základním stavebním prvkem je u rastrové struktury tzv. **buňka** (cell, pixel).
- Buňky jsou organizovány do **mozaiky**.
- Jednotlivé buňky obsahují **hodnoty** (values).
- **Typy tvarů buněk:**
  - čtvercová buňka (lattice, grid)
  - trojúhelníková buňka,
  - hexagonální buňka.

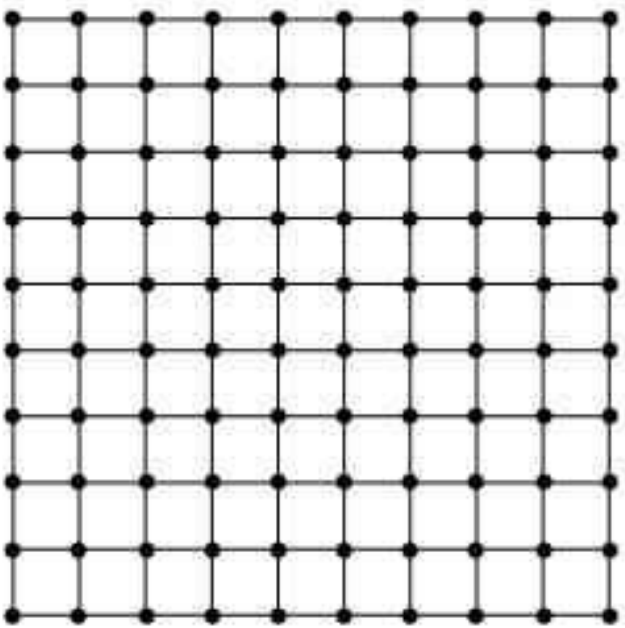


# Typy rastrové reprezentace

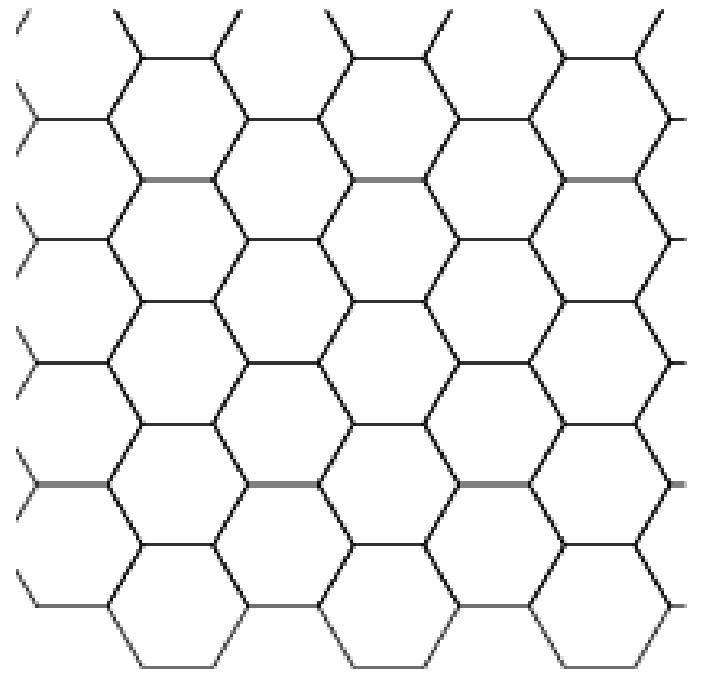
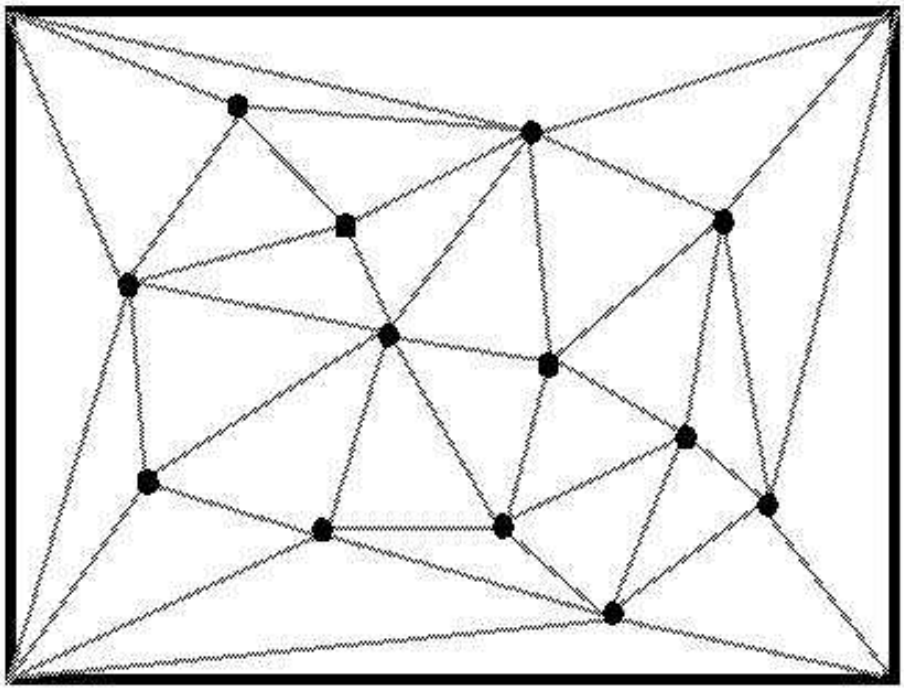
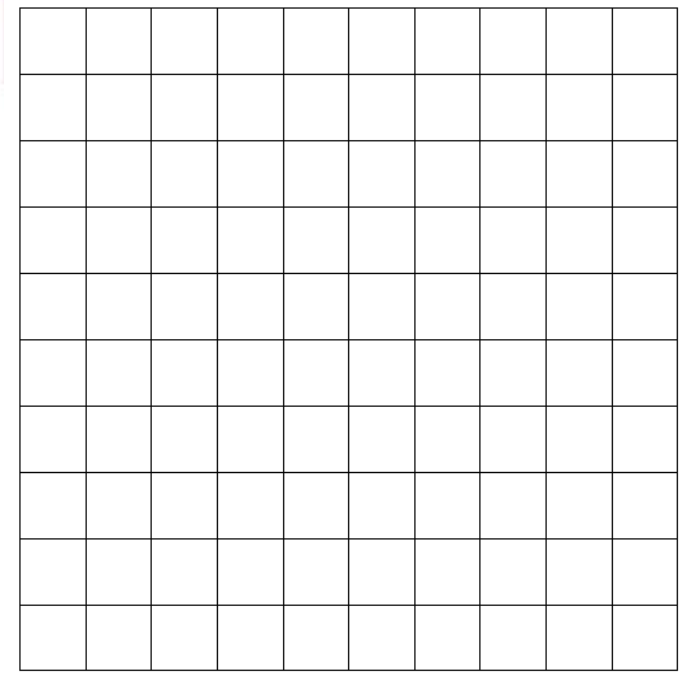
**Rastrovou reprezentaci můžeme rozlišit podle způsobu dělení prostoru na:**

- **pravidelné (regular) - všechny buňky mají stejnou velikost a tvar.**
  - jednodušší pro ukládání a zpracování údajů, zabírají ovšem na disku mnoho místa.
- **nepravidelné (irregular) - velikost i tvar jednotlivých buněk se liší.**
  - mohou mnohem lépe reprezentovat danou lokalitu (příklad roviny + zvlněná krajina),
  - zpracovávání je algoritmicky i výpočetně náročné. Hlavně pro DMR.

Lattice Network



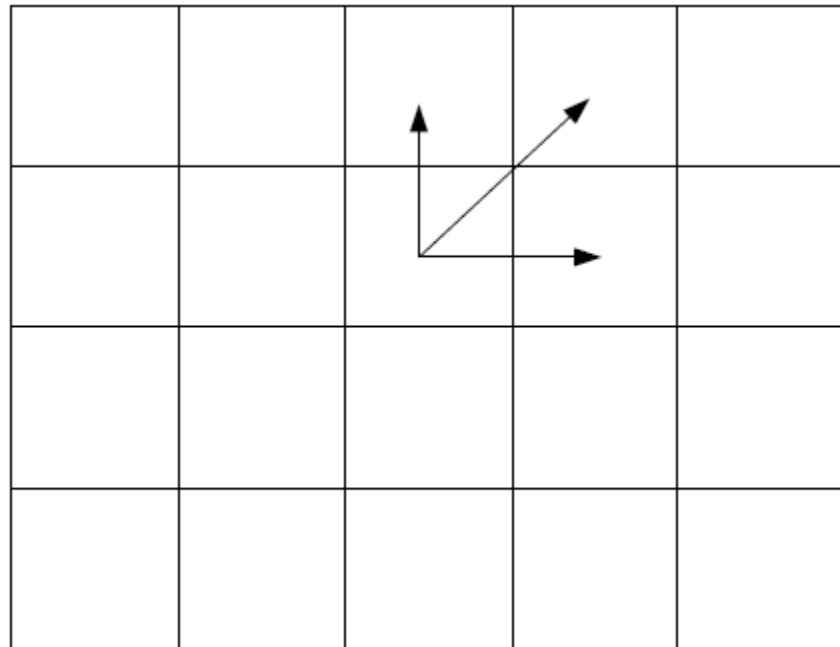
10 x 10 Grid



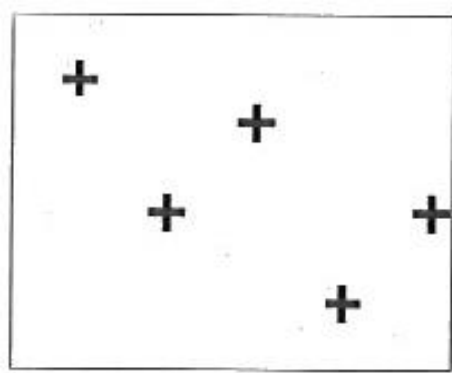


# Topologie v rastru

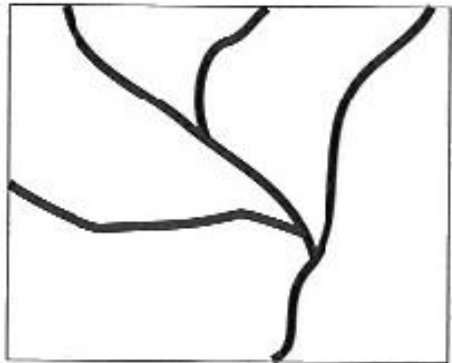
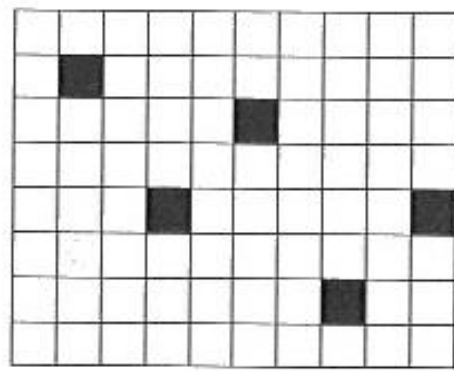
- **Topologie je v rastrovém modelu definována implicitně (je jasné, kdo je čí soused), tudíž není nutné ji explicitně ukládat jako pro vektorový model!**



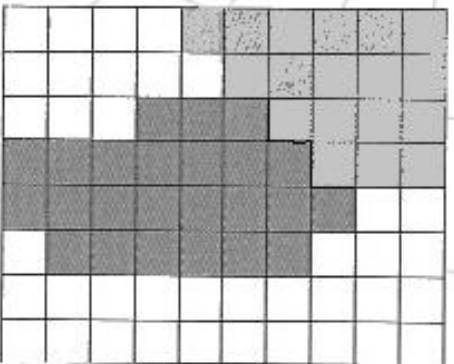
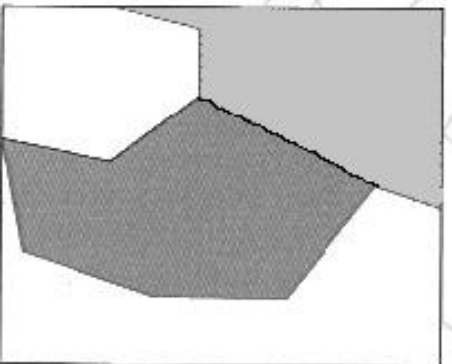
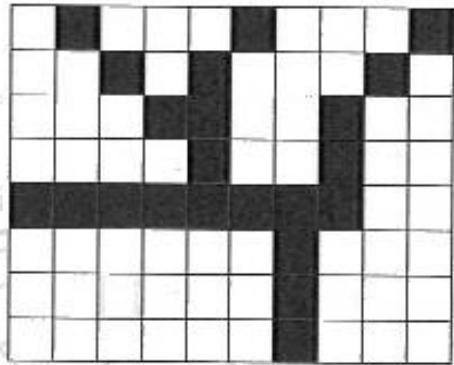
# Reprezentace geometrie v rastru



*Point features represented in a grid.*



*Linear features represented in a grid.*



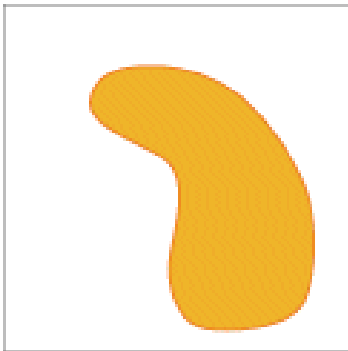
- **rastrová datová struktura může nést informace o bodech, liniích a plochách.**
- **Odlišné možnosti převodu mezi vektorem a rastrem.**



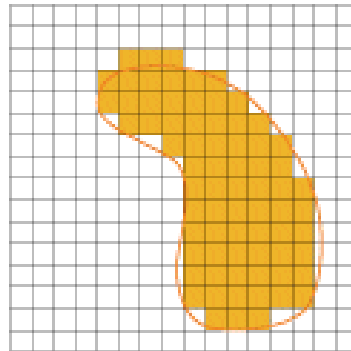
# Faktory ovlivňující vyjádření v rastru - rozlišení

- Vliv velikosti buňky ( $\sim$  rozlišení) na tvar objektů (+ a -)

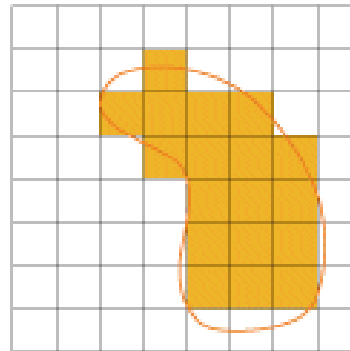
71 m<sup>2</sup>  
polygon



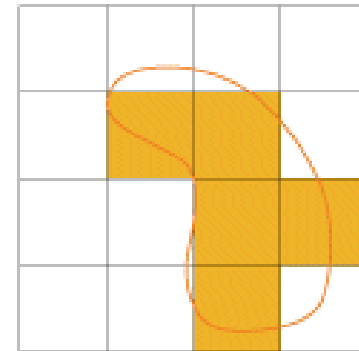
73 m<sup>2</sup>  
1 m cell  
16 x 16 cells



72 m<sup>2</sup>  
2 m cell  
8 x 8 cells



80 m<sup>2</sup>  
4 m cell  
4 x 4 cells



**PRO**

- Smaller cell size
- Higher resolution
- Higher feature spatial accuracy

**PROTI**

- Larger cell size
- Lower resolution
- Lower feature spatial accuracy

**PROTI**

- Slower display
- Slower processing
- Larger file size

**PRO**

- Faster display
- Faster processing
- Smaller file size



# Faktory ovlivňující vyjádření v rastru

**datové rozlišení** („barevná hloubka“ rastru) - popisuje počet bitů použitých k popisu určité barvy pixelu v bitmapovém obrázku :

- **binární rastr** (0x1, výskyt x nevýskyt) – záznam 1 bitem.
- **8bitový rastr ( $2^8$ )** – 256 různých celočíselných hodnot, záznam 1 bajtem.
- **24bitový rastr** – 1,6 milionu různých celočíselných hodnot, 3 bajty.
- **kontinuální rastr** – hodnoty v reálných číslech, záznam 4 nebo 6 bajty.

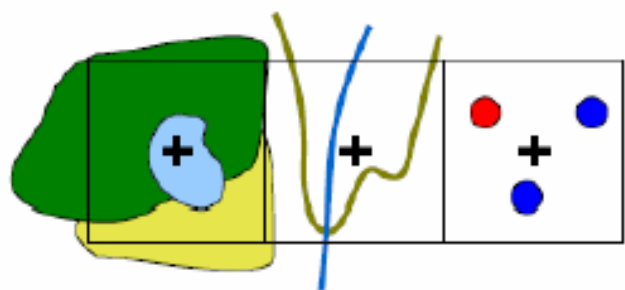
**způsob přiřazení hodnot zobrazovaného atributu (kvantitativní data) – při tvorbě modelu:**

- jako bodová hodnota změřená **kdekoli v ploše buňky**
- jako **aritmetický průměr** u několika bodových měření
- jako **vážený aritmetický průměr**, kde váhou je plošný rozsah jednotlivých hodnot
- jako **maximální nebo minimální hodnota atributu** v ploše buňky
- jako hodnota atributu **s největší váhou** (i pro kvalitativní).

# Řešení konfliktů

Problém - **jedna výsledná buňka obsahuje více různých objektů**. Pro řešení této se používají 3 základní metody, z čehož **první dvě se používají pro převod bodů, linií i polygonů** a zbývající jen pro **převod polygonů**:

- **Metoda dominantního typu** vychází z principu, že u buňky, do které zasahuje více objektů, se vyjádří podíl její plochy, zabíraný každým z objektů a hodnota objektu s největším podílem je pak buňce přiřazena (u bodů a linií se podíl plochy nahrazuje počtem a příp. délkou objektů, které buňka obsahuje).
- **Metoda nejdůležitějšího typu** buňce přiřadí hodnotu, která je považovaná za nejdůležitější z hlediska aplikace.
- **Metoda centroidu**, buňka má přiřazenou hodnotu definovanou polohou jejího středu při průmětu do vektorové reprezentace.



Dominantní typ



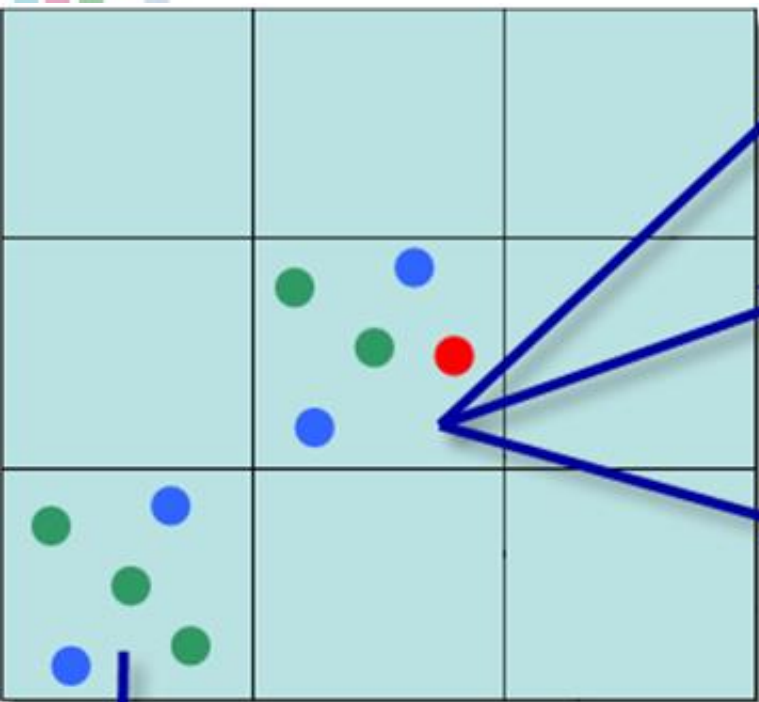
Nejdůležitější typ



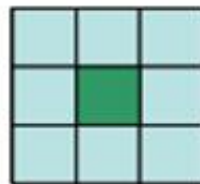
Centroidy



# Příklady – pravidla pro rasterizaci bodů v ArcGIS

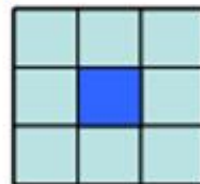


FID	Attribute
1	Green
2	Red
3	Blue
4	Blue
5	Green



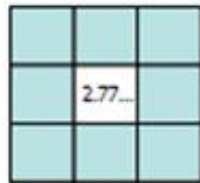
Field = Attribute  
 Method = MOST\_FREQUENT  
 Priority = NONE  
 Outcome = Green  
 Reason = Lowest FID

FID	Attribute	PriorityFID
1	Green	1
2	Red	1
3	Blue	1
4	Blue	3
5	Green	2



Field = Attribute  
 Method = MOST\_FREQUENT  
 Priority = PriorityFID  
 Outcome = Blue  
 Reason = Highest priority

FID	ValueFID
1	1
2	8
3	5
4	3
5	2

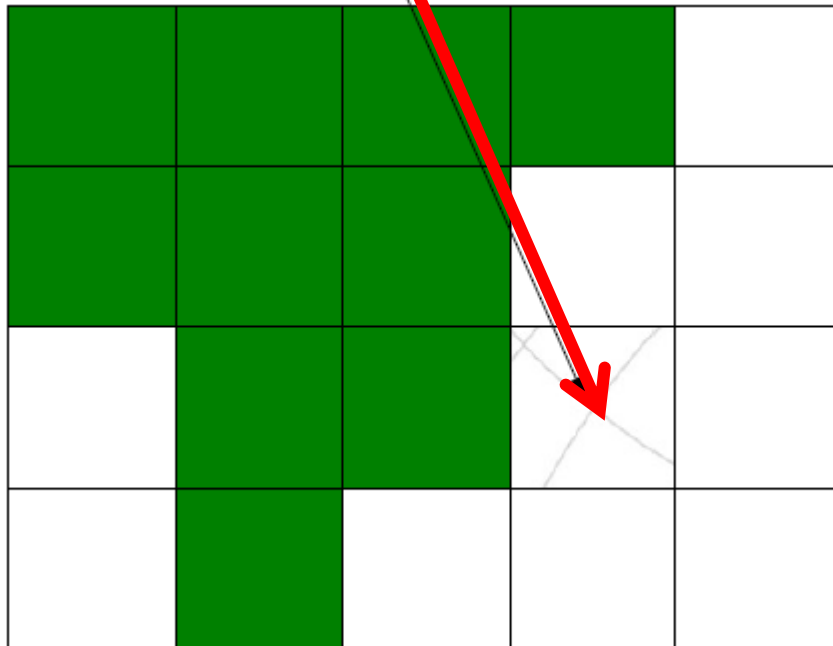


Field = ValueFID  
 Method = STANDARD\_DEVIATION  
 Priority = Ignored  
 Outcome = 2.774887323379517  
 Reason = Priority field is only used with MOST\_FREQUENT



# Prázdné buňky

- Pokud je hodnota buňky definována jako prázdná (NoData), znamená to, že tato **buňka nese žádnou informaci o prostoru**, který reprezentuje.
- **0 je validní hodnota!**
- 999 obvykle použito pro No data



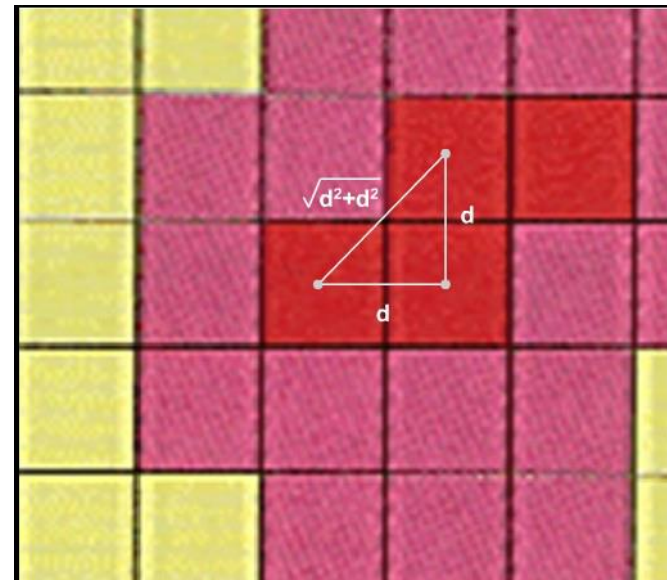
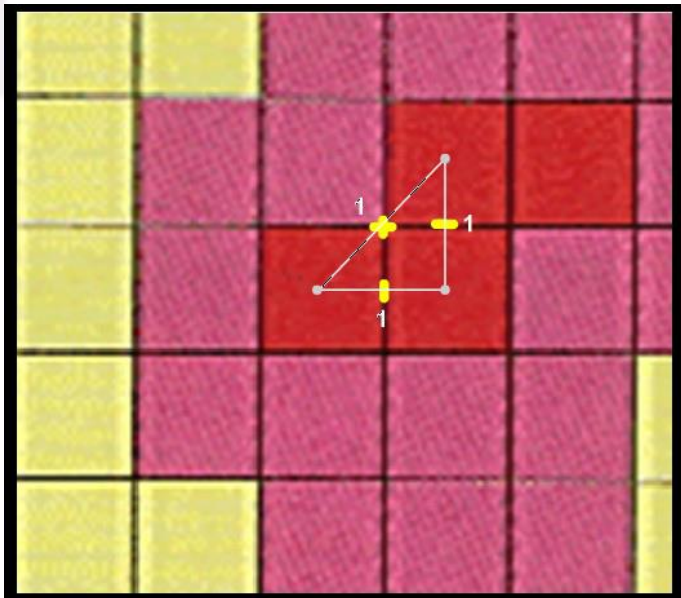




# Metrika čtvercové mřížky

- V geometrii nastává problém metriky (způsob definice vzdálenosti dvou buněk) – odlišná vzdálenost středu čtverců.
- Euklidovská metrika

$$d = \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2}$$





# Rastrová data výhody a nevýhody

## výhody

- jednoduchost datové struktury
- snadné překrývání a kombinace obrazů s různým obsahem
- rychlé dotazování
- snadná tvorba uživatelských nadstaveb
- jednoduchá kombinace s jinými daty rastrové povahy (DPZ)
- snadné provádění analytických operací

## nevýhody

- **značná paměťová náročnost (velký objem dat)**
- omezená přesnost, daná rozlišením rastru a orientací rastru (výpočty délek, vzdáleností, ploch ...)
- kvalita výstupů závislá na rozlišení rastru (nižší vizuální kvalita rastrových výstupů)
- nevhodnost pro síťové analýzy



# Kompresní techniky pro rastry

- **Ztrátové**

- komprimují lépe než neztrátové
- dochází ke ztrátě informace => někdy nevhodné!

- **Neztrátové**

- Run Length Codes – RLC
- Run Length Encoding – RLE
- Čtyřstrom – QuadTree
- Adaptivní komprese

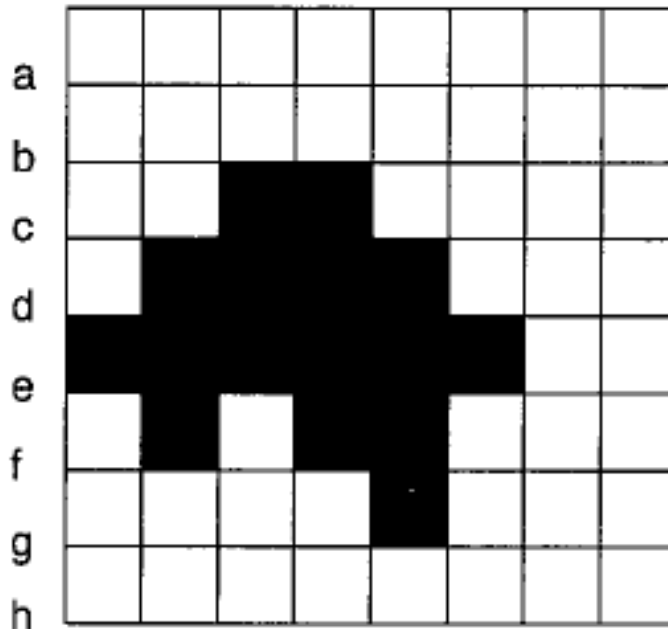


# Run Length Codes

- Definuje příslušnost buněk rastru k objektu po řádcích nebo sloupcích, přičemž udává jen **začátek a konec úseku buněk** v řádku či sloupci.
- Pro černobílé/binární rastry

row #	column # run1 begin	column # run1 end	column # run2 begin	column # run2 end
-------	------------------------	----------------------	------------------------	----------------------

1 2 3 4 5 6 7 8



Řádek c 3,4  
Řádek d 2,5  
Řádek e 1,6  
Řádek f 2,2 4,5  
Řádek g 5,5

# Run Length Encoding

- Využití maticového zápisu dat.
- Efektivní při rozsáhlých homogenních oblastech dat

**1 1 1 1 5 5 9 9 9 9 9 9 9 2 9 9 9**  
**(4 1)(2 5)(7 9)(1 2)(3 9)**

- Heterogenní ☹

**0 1 0 1 2 3 5 2 1 4**

**(1 0)(1 1)(1 0)(1 1)(1 2)(1 3)(1 5)(1 2)(1 1)(1 4)**

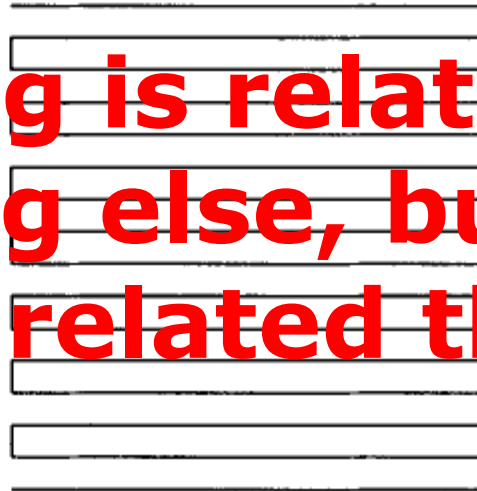
**Jak zefektivnit kompresi?**

# Způsob procházení rastru

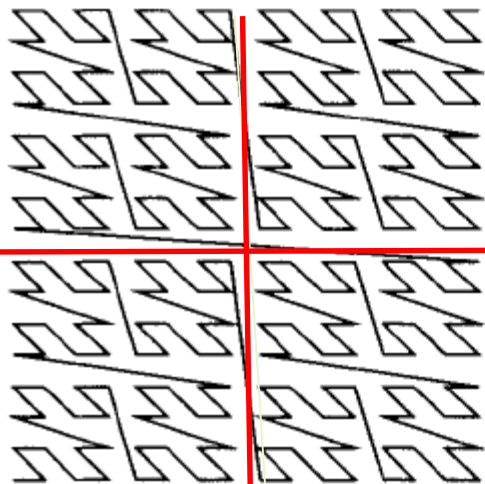
**Everything is related with everything else, but near things are more related than distant things.**



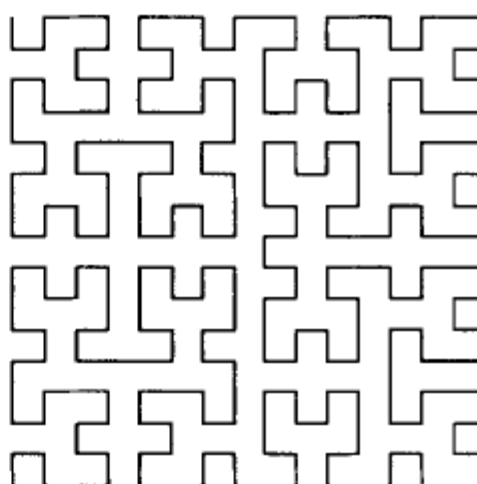
(a) Row Order



(b) Row-prime Order



(c) Morton Order

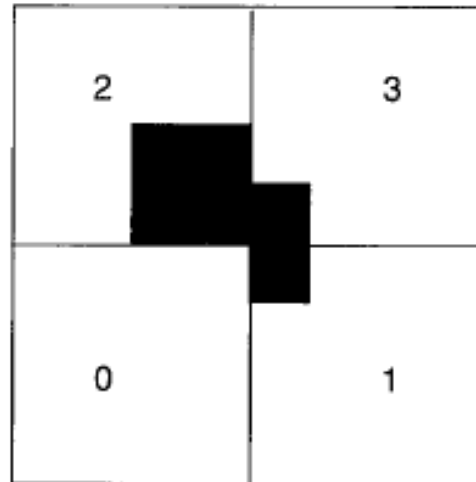
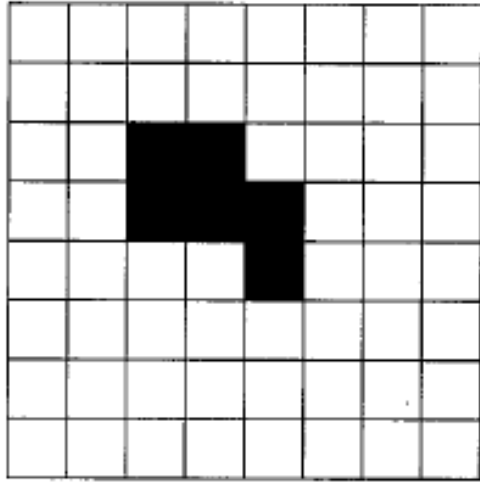


(d) Pi-Order

- A) a B) - alternativy postupného procházení.
- C) a D) - prostor vyplňující křivky (space-fill).
- B) a D) jsou více efektivní - souvislost s Toblerovým zákonem (First law of Geography, autokorelace).



# Quad tree - čtyřstrom



- Hierarchické uložení
- Dělení kvadrantů až do doby, kdy jsou homogenní.

