

Detekce, prokazatelnost a vizualizace extrémů demografických dat ve statistických souborech

VOŽENÍLEK Vít, KAŇOK Jaromír
&
TUČEK Pavel



UNIVERZITA PALACKÉHO V OLOMOUCI
KATEDRA GEOINFORMATIKY
TR. SVOBODY 26, OLMOUC



extrémní sport



extrémní sucho



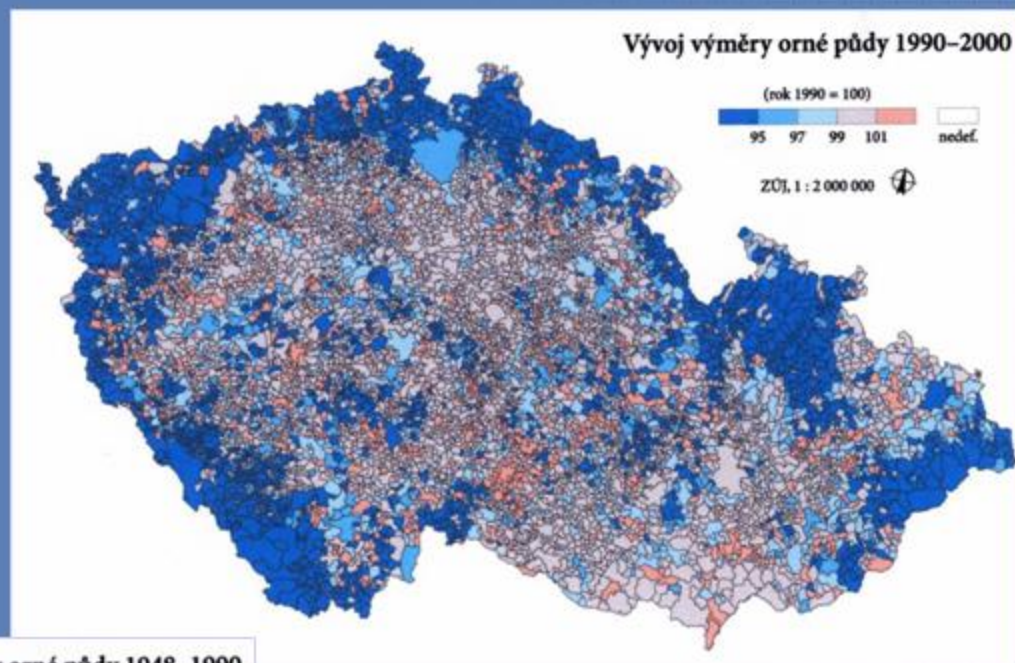
extrémně krátká sukně



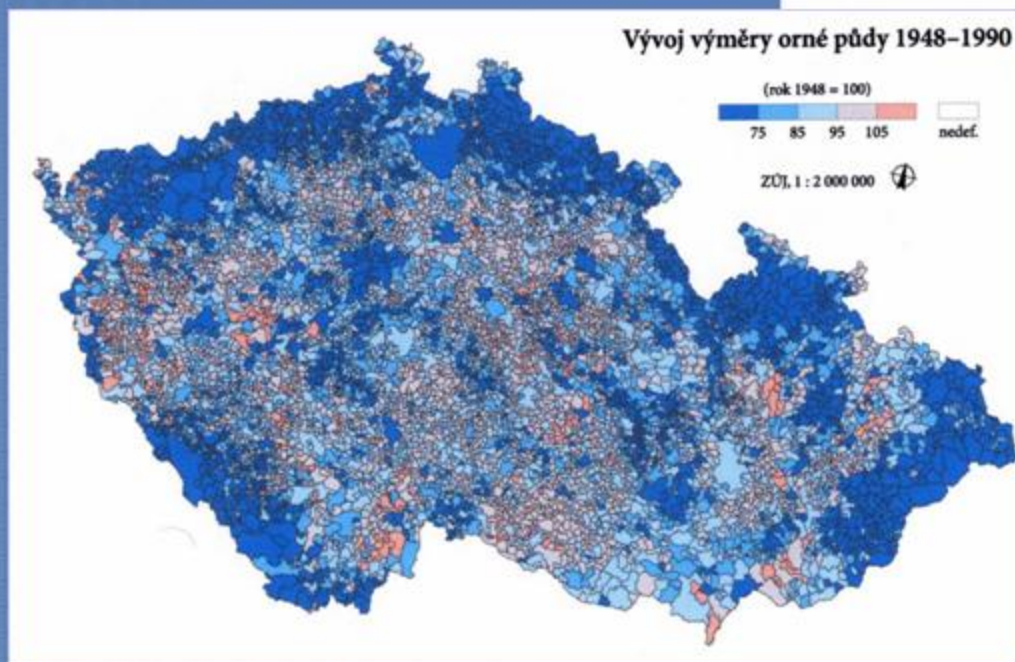
extrémní řešení

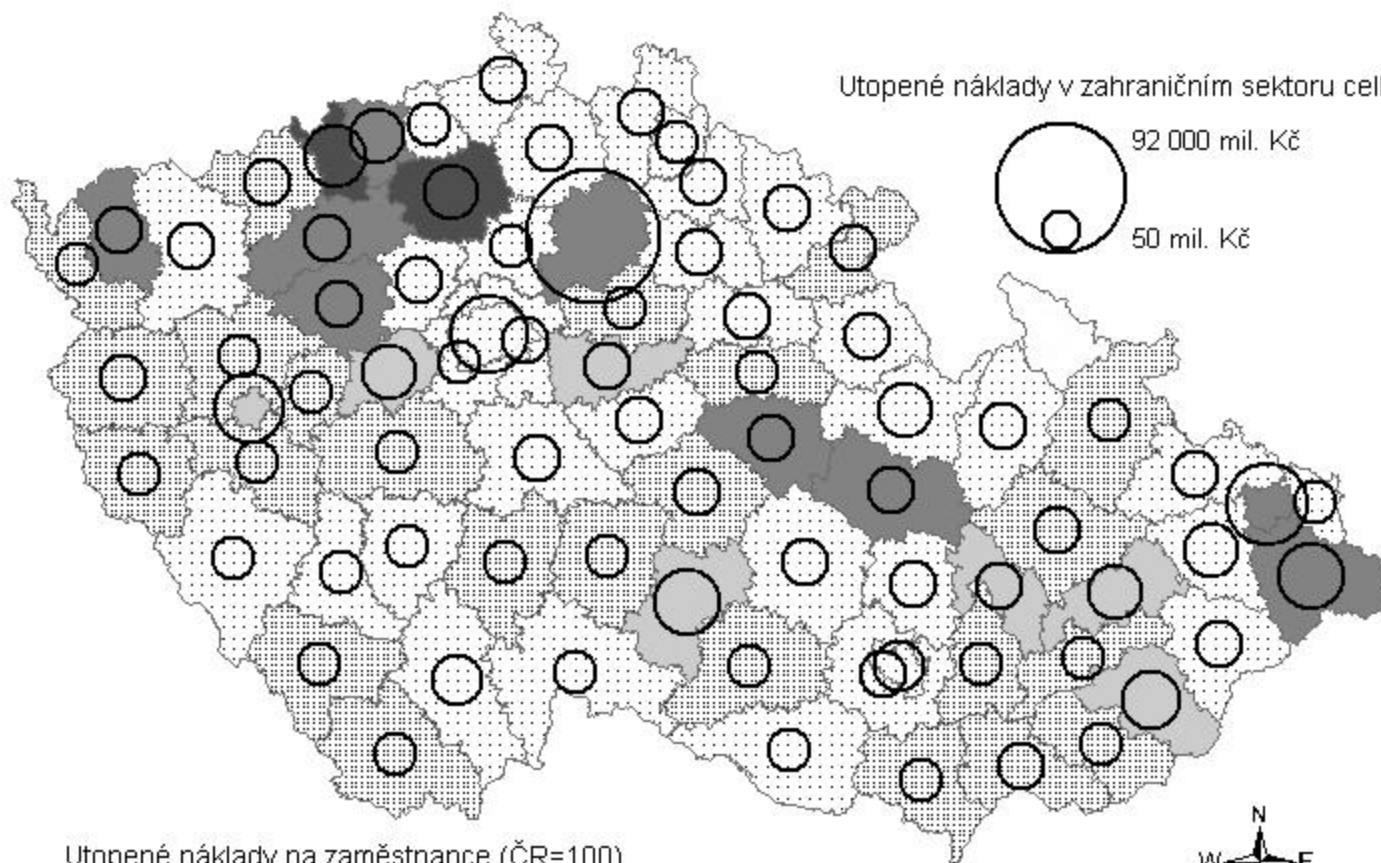


Vývoj výměry orné půdy 1990–2000



Vývoj výměry orné půdy 1948–1990



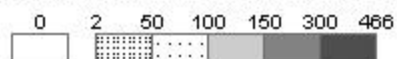


Utopené náklady v zahraničním sektoru celkem

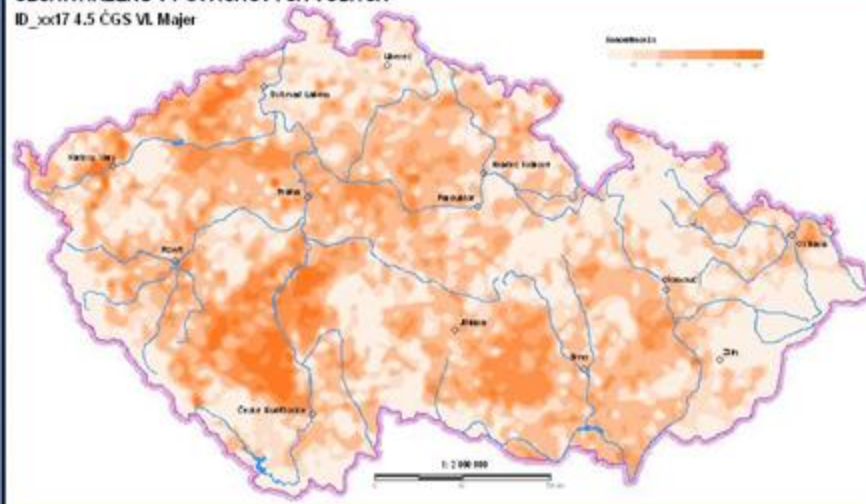
92 000 mil. Kč

50 mil. Kč

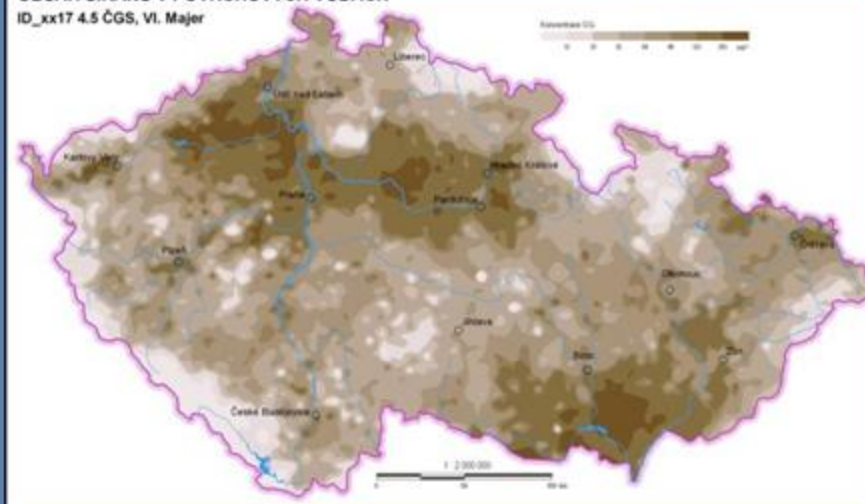
Utopené náklady na zaměstnance (ČR=100)



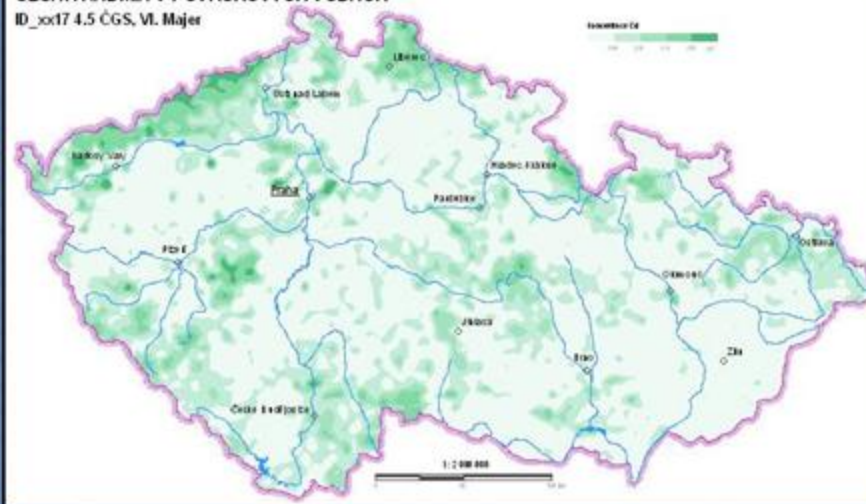
ID: xx17 4.5 ĆGS VL Majer



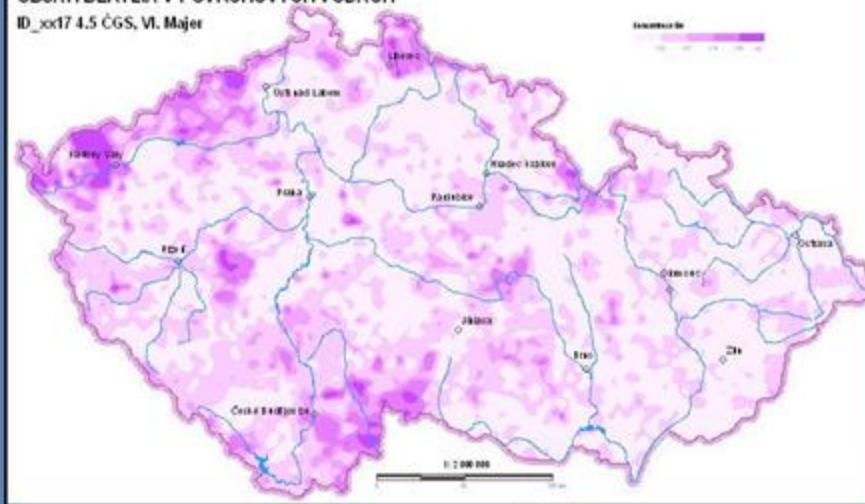
ID_xx17 4.5 ČGS, VI. Majer



ID_0017 4.5 ĆGS, VI. Majer

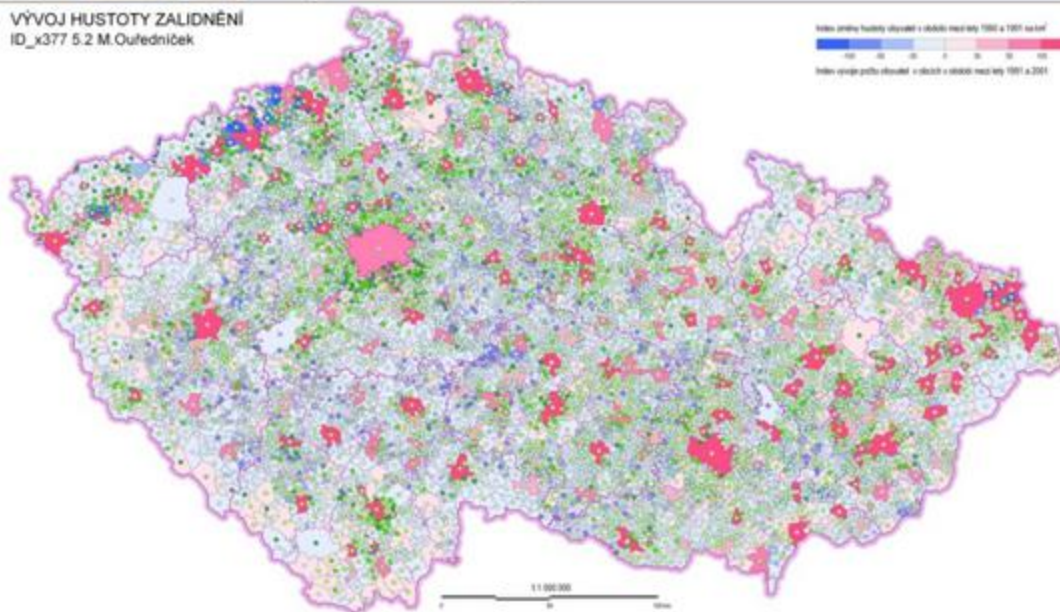


ID: xx17 4,5 ĆGS, VI, Majer



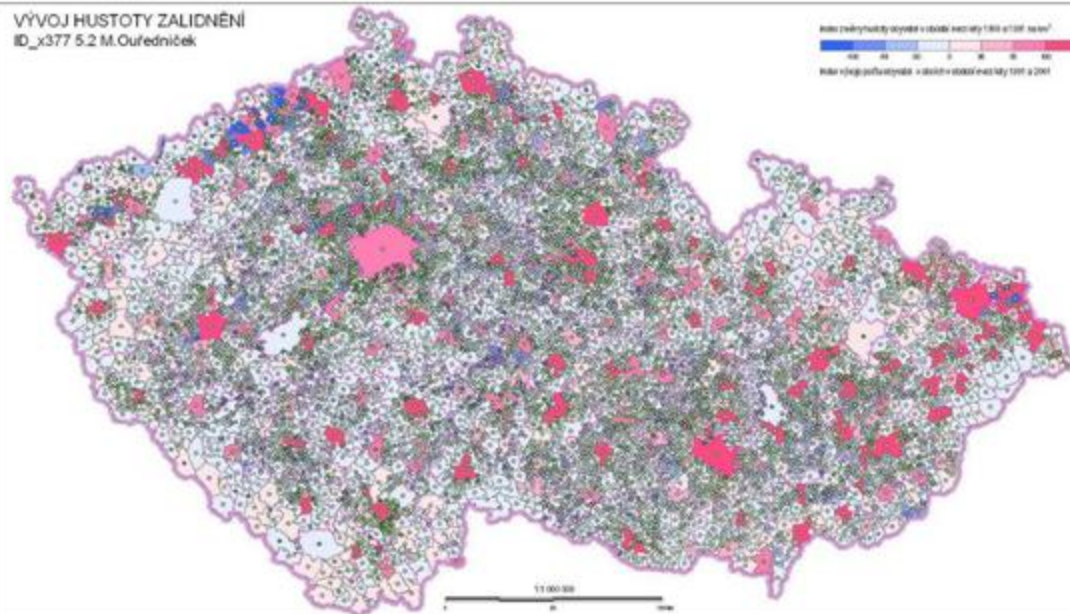
VÝVOJ HUSTOTY ZALIDNĚNÍ
ID_x377 5.2 M.Oufedníček

Index změny hustoty výsadby v období mezi lety 1991 a 2001 na km²
Index výsadby prvních výsadby v období mezi lety 1991 a 2001



VÝVOJ HUSTOTY ZALIDNĚNÍ
ID_x377 5.2 M.Oufedníček

Index změny hustoty výsadby v období mezi lety 1991 a 2001 na km²
Index výsadby prvních výsadby v období mezi lety 1991 a 2001

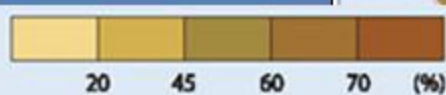


VLASTNICKÁ STRUKTURA ZEMĚLSKÝCH PODNIKŮ

XXX
Ivan Bilič, Vít Jančák
1 : 2 000 000
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

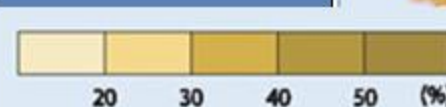
PODÍL DRUŽSTEV

XXX



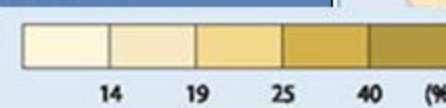
PODÍL OBCHODNÍCH SPOLEČNOSTÍ

XXX



PODÍL SOUKROMÉ HOSPODÁŘICÍCH ROLNÍKŮ

XXX



extrém

- největší/nejmenší hodnota

Extrém = krajnost, výstřednost, mimořádnost.



- Konkrétní význam může být:
- Extrém funkce – matematický pojem
- Extremismus
- Extrémní programování (XP) – metoda vývoje SW
- Extrémní sporty

Statistika

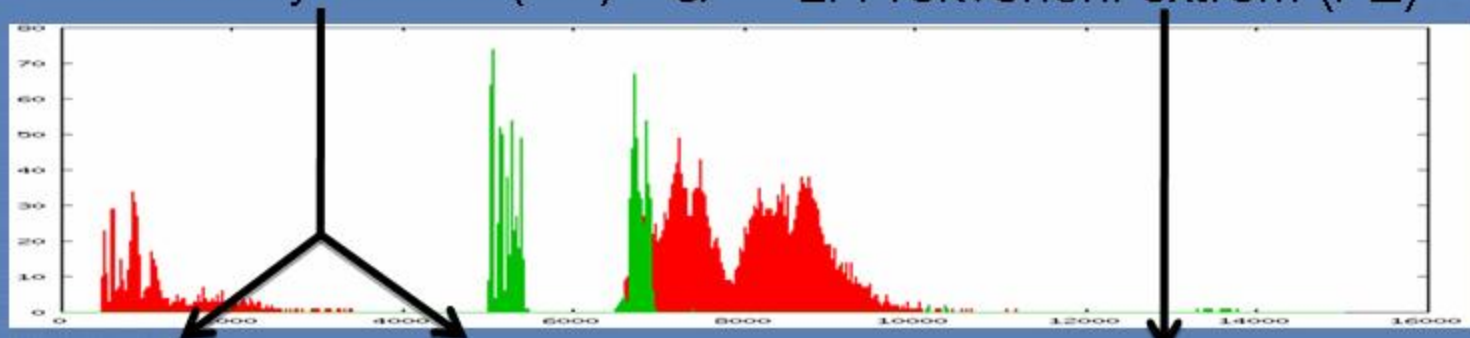
- zkoumaný jev – statistická náhodná veličina

detekce → prokazování → interpretace → vizualizace

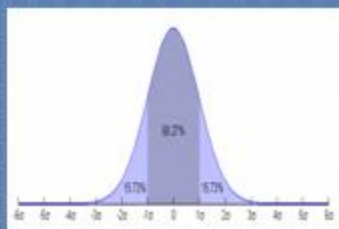
EXTRÉMY Z POHLEDU MATEMATIKA A STATISTIKA

Extrém = obecně malá nebo velká hodnota

1. Datový extrém (DE) & 2. Frekvenční extrém (FE)



Teoretická mez X Naměřená data



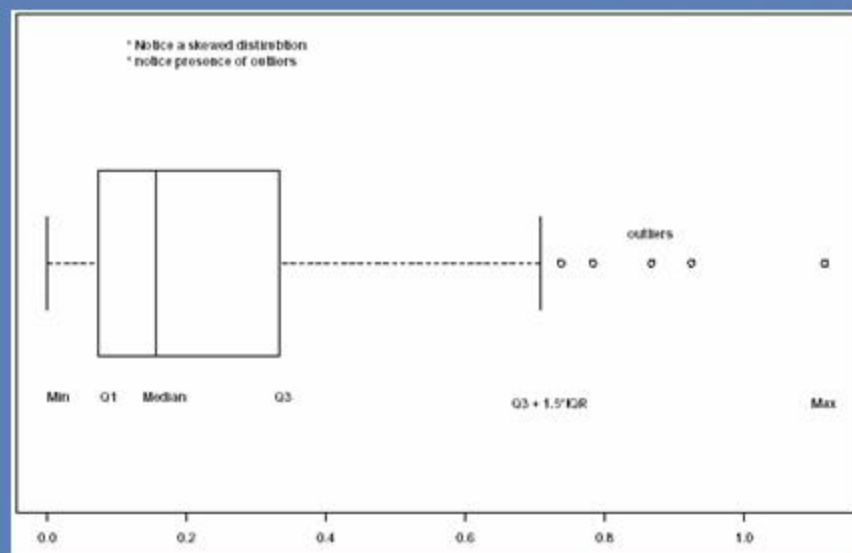
$$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

Min, Max, Boxplot...

FE = data s velkou nebo malou četností výskytu bez ohledu na jejich hodnotu

! Pozor na to, co mapa znázorňuje !

OUTLIER



Pozorování x_i se nazývá outlierem, jestliže se jedná o nápadně vybočující pozorování způsobené chybným měřením, chybným zápisem nebo o signifikantně vybočující a nesouvisející datový údaj

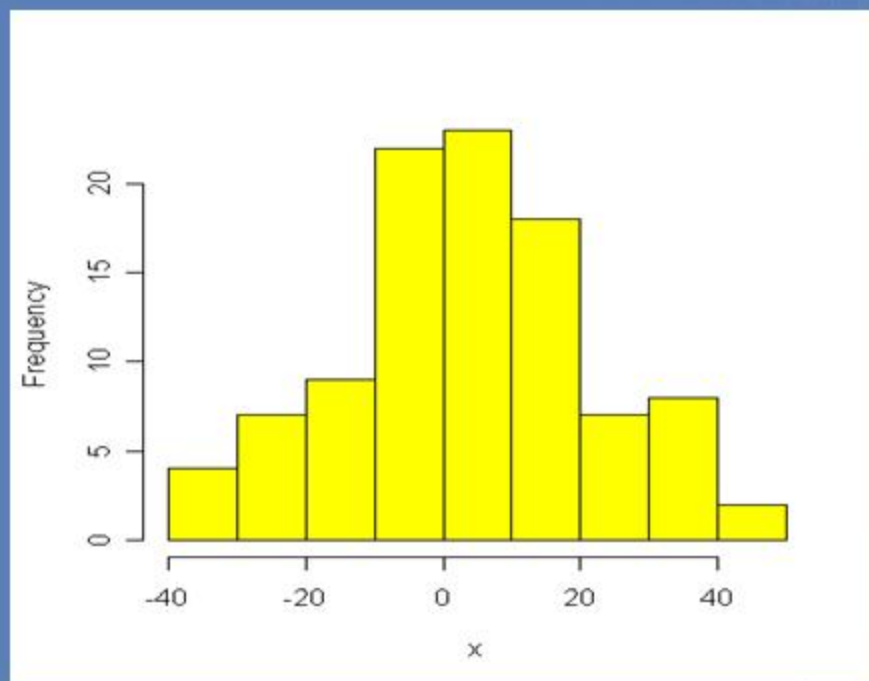
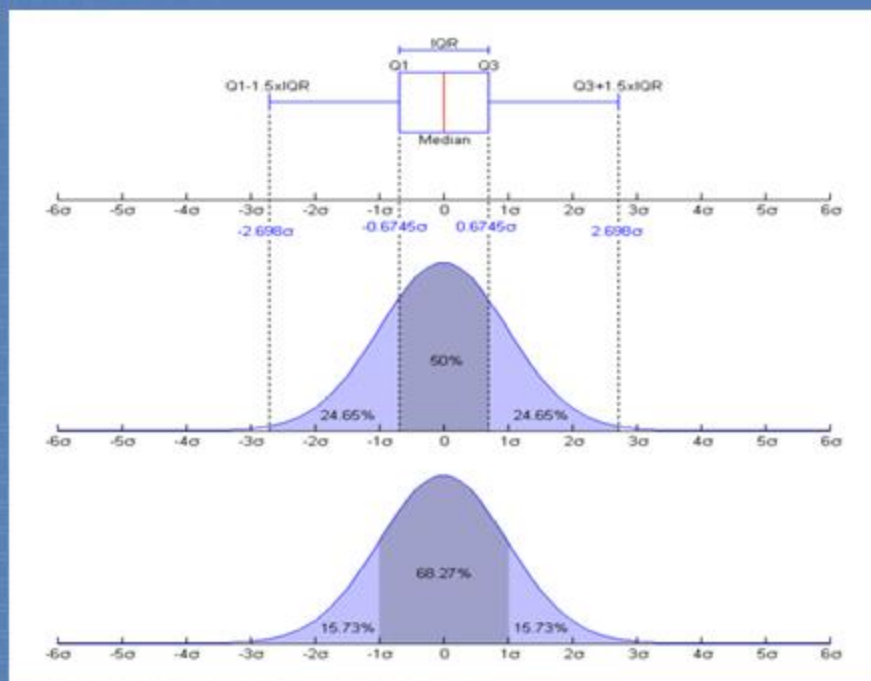
Zpracování dat



BOXPLOT – HISTOGRAM

Jednoduché a přitom velice sofistikované metody odhalení existence extrémů v datových sadách a teoretických modelech je založena na analýze pravděpodobnostního rozložení jednotlivých hodnot.

Jak již bylo řečeno, extrémy se nacházejí jak v datové tak ve frekvenční oblasti datové sady. Je tedy více než zřejmé že podrobnou analýzou těchto dvou oblastí dosáhneme identifikace extrémních hodnot.





VÝPOČET OPTIMÁLNÍHO POČTU INTERVALU

Metoda postupného dělení intervalů

- Metoda založená na využití míry informační entropie. Při tvorbě map hodnotíme míru zachované informace při postupné generalizaci

$$H(Jev) = - \sum_{i=1}^n P(x_i) \log_2 P(x_i)$$

- Využíváme postupného dělení intervalů a výpočtu entropie v poměru ku počtu zanechaných intervalů. Tento algoritmus vede k navržení optimálního počtu intervalů, které při tvorbě mapy použít.

$$\frac{H(Jev)}{N(int)} \rightarrow \max \quad \vee \quad \text{Relativní entropie} \rightarrow \max$$

- V okamžiku, kdy známe optimální počet intervalů, pak nastává situace, kdy je třeba rozhodnout o jejich poloze a v tom nám pomůže znalost statistického chování zkoumaného jevu a hlavně hranice, od které považujeme hodnoty za extrémní.




POJMY

- Extrémní hodnoty v datové oblasti

- a) Datová extrémní hodnota – zjištěná, změřená
- b) Datová extrémní hodnota – teoretická, modelová
- c) Nejvyšší zjištěná, nebo naměřená hodnota

- Extrémní hodnoty ve frekvenční oblasti

- a) frekvenční extrémy skutečné
 - b) frekvenční extrémy teoretické
- 



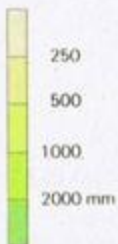
Najdeme všechny druhy extrémů na současných mapách ?

- Hledejme odpověď

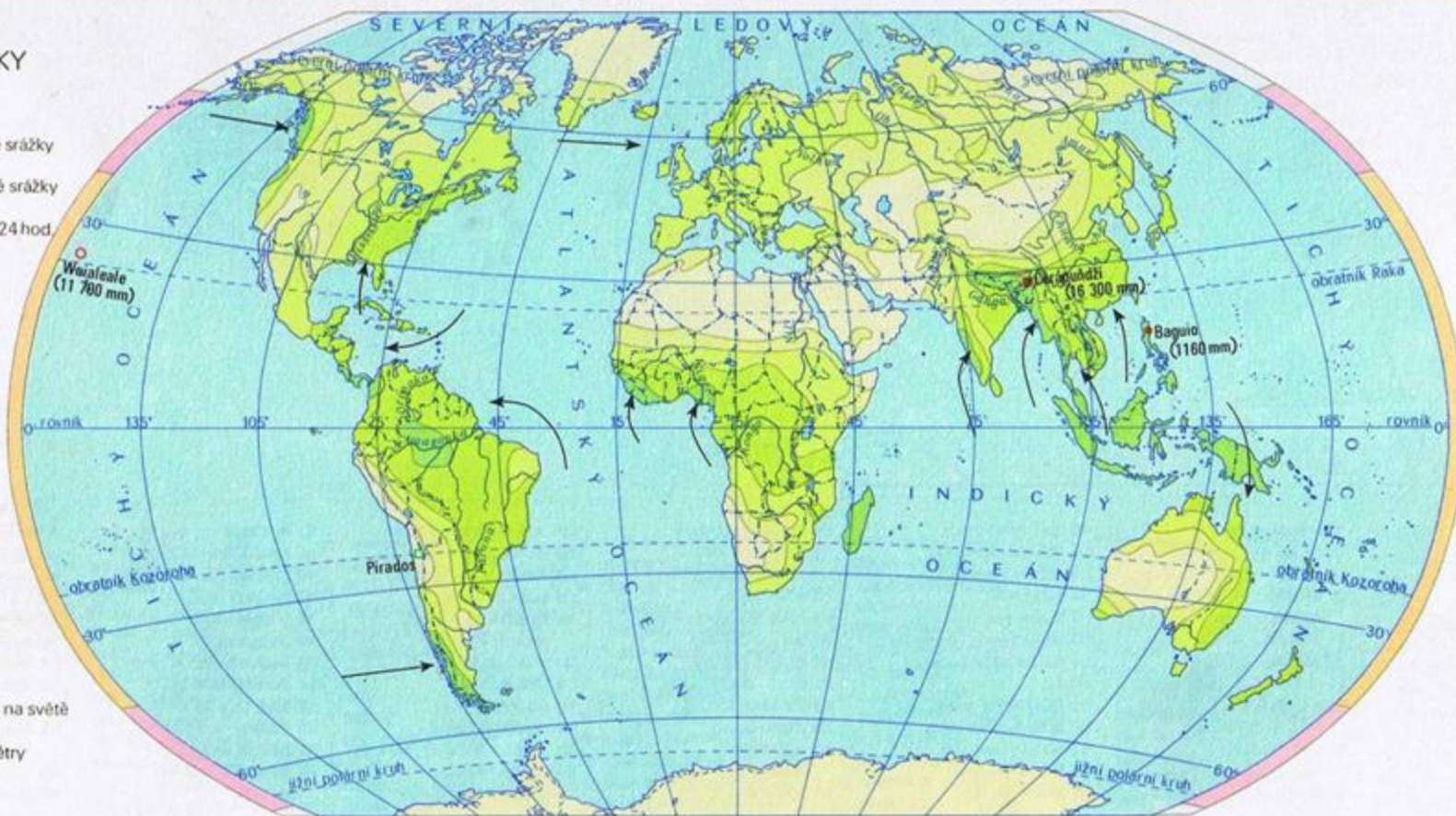
- a) Extrémní hodnota – zjištěná, změřená ?
b) Nejvyšší zjištěná nebo naměřená hodnota ?

ROČNÍ SRÁŽKY

- největší průměrné srážky
- největší naměřené srážky
- největší srážky za 24 hod.

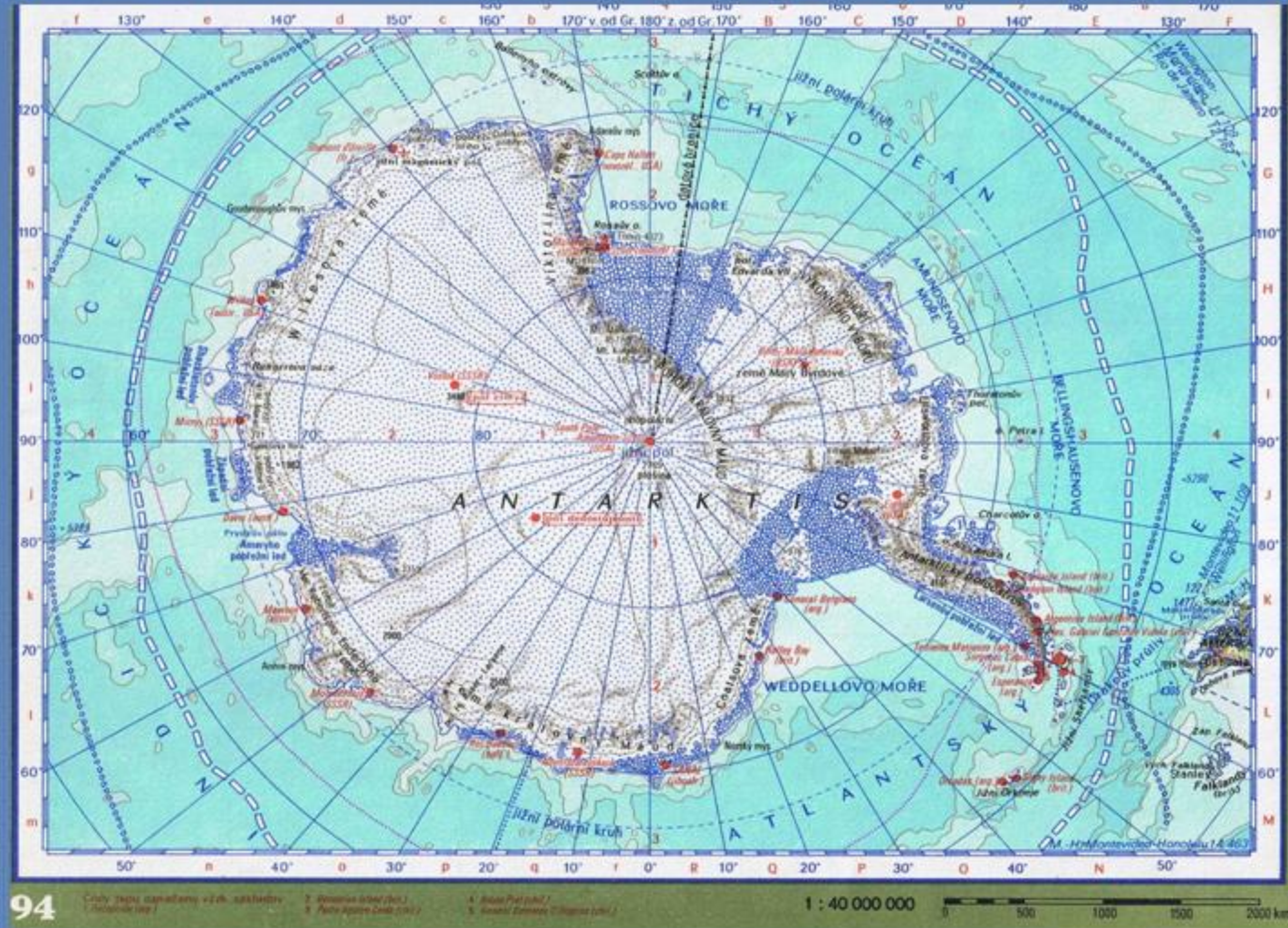


- nejsušší místo na světě
- vláhonosné větry

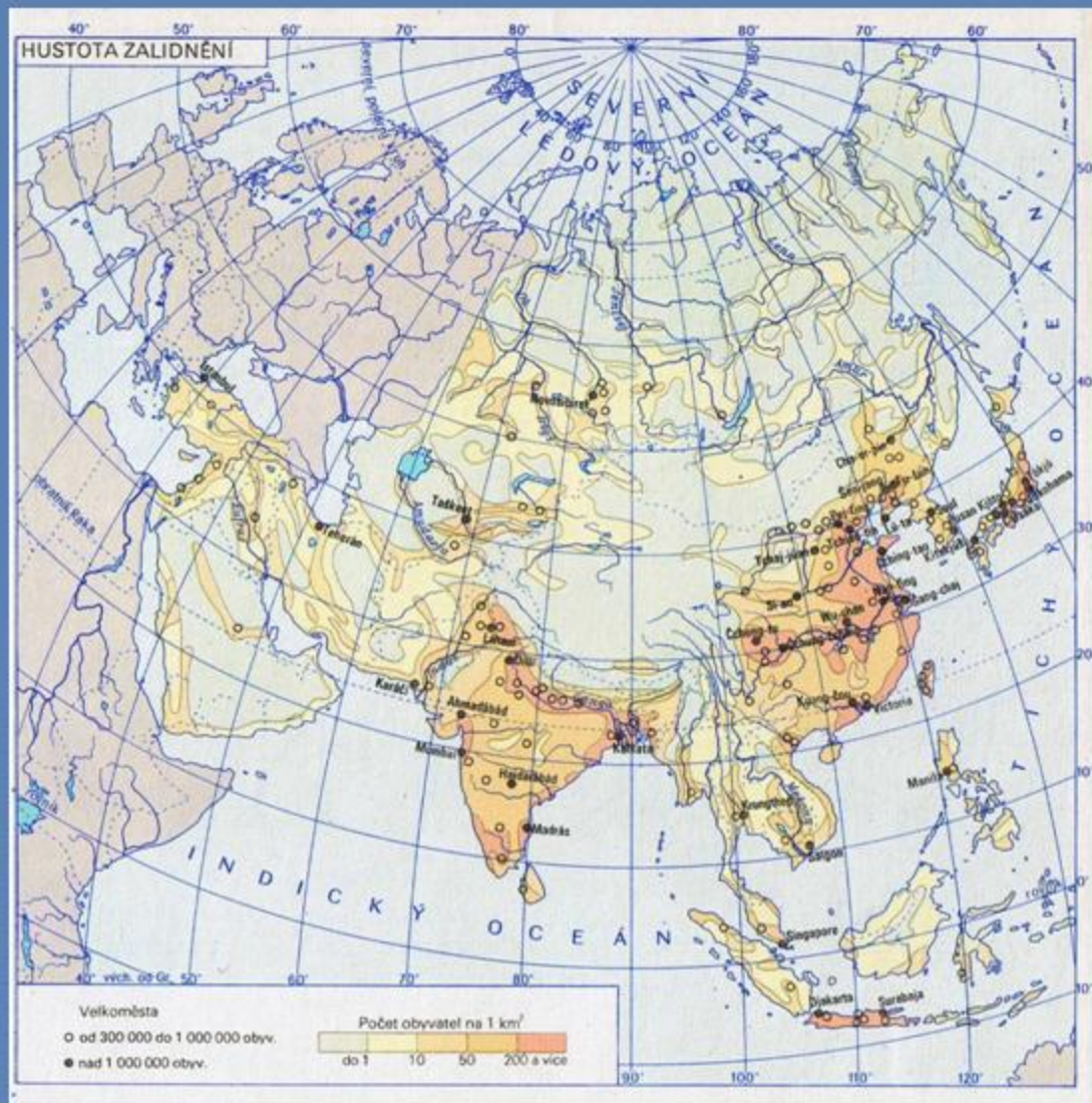


1 : 200 000 000

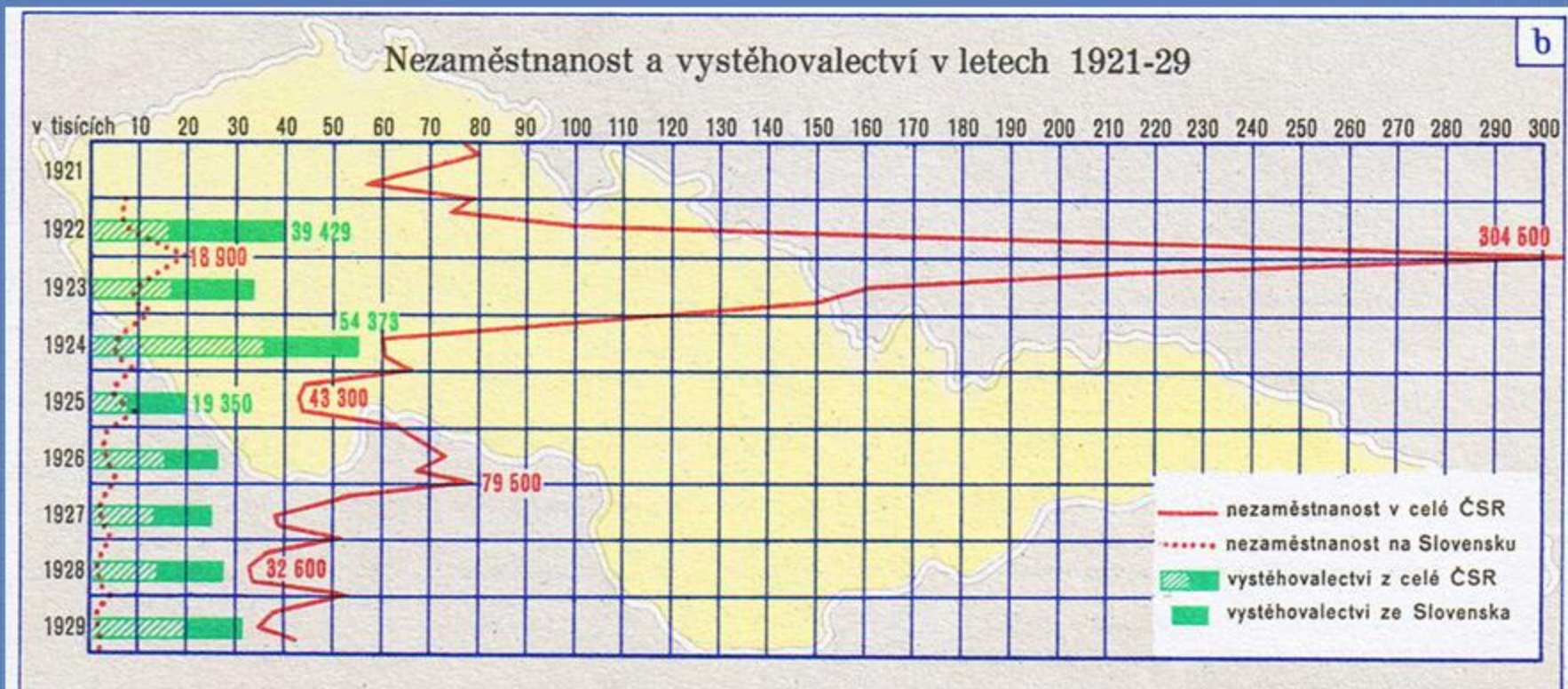
- a) Extrémní hodnota – zjištěná, změřená
 b) Nejvyšší zjištěná nebo naměřená
 hodnota ?



- a) Extrémní hodnota – zjištěná, změřená?
b) Nejvyšší zjištěná nebo naměřená hodnota?



Nejvyšší zjištěná hodnota (v grafu)

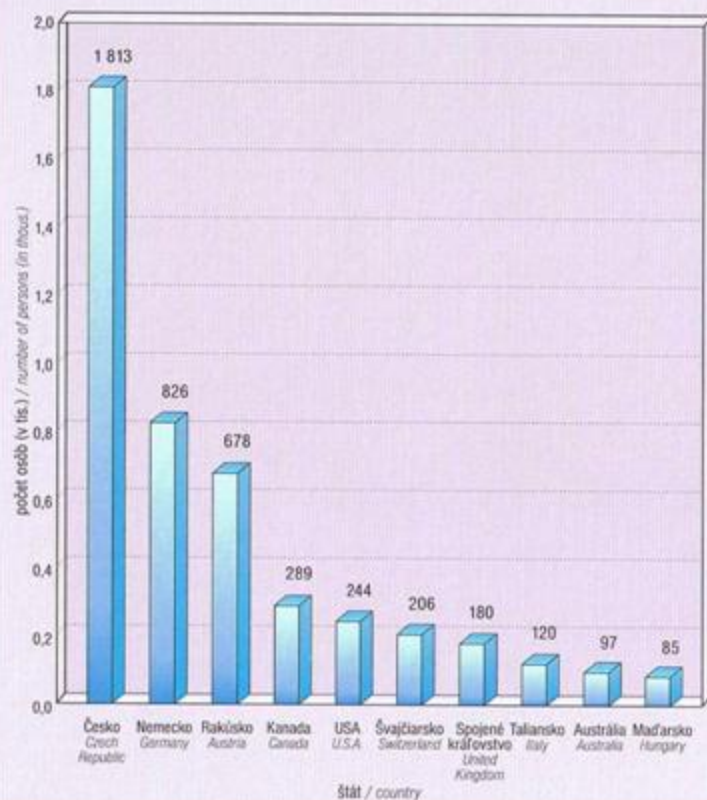


Nejvyšší zjištěná hodnota (v diagramu)

Vystáhovaní zo Slovenska v rokoch 1999 – 2003 podľa najvýznamnejších krajín

Emigrants from Slovakia by the most significant
countries, 1999 – 2003

Boris Divinský



Nejvyšší zjištěná hodnota (v kartodiagramu)

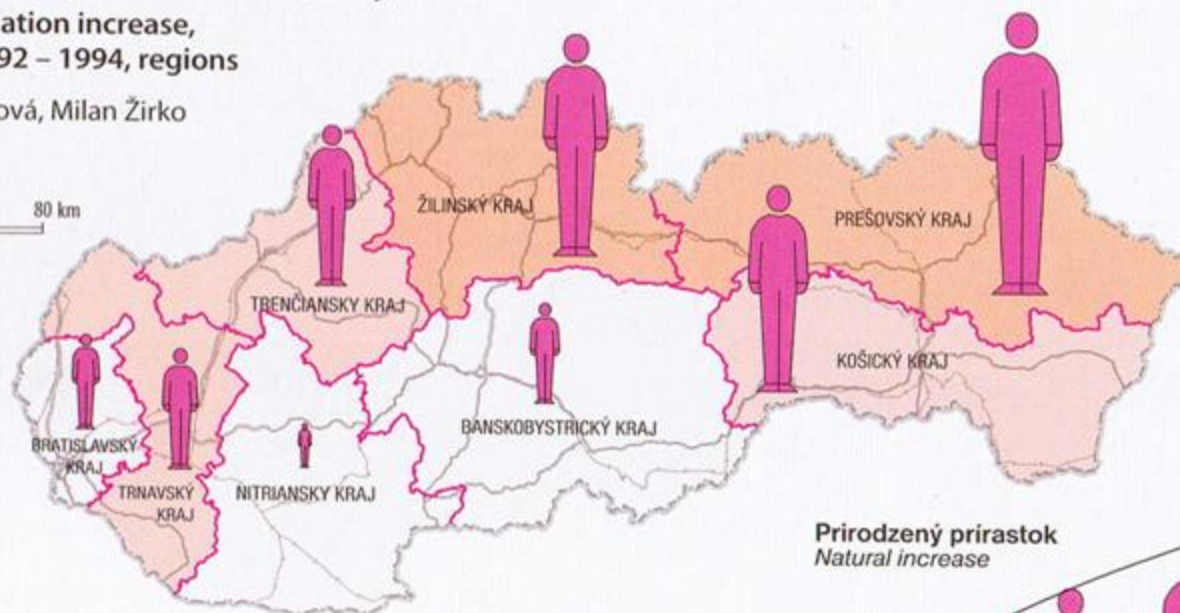
Prirodzený prírastok obyvateľstva priemer za roky 1992 – 1994, kraje

Natural population increase,
average of 1992 – 1994, regions

Jana Marenčáková, Milan Žirko

1 : 4 000 000

0 40 80 km

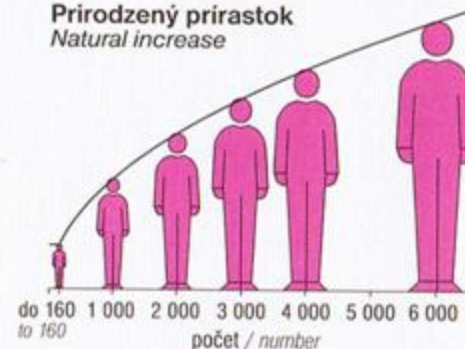


Prirodzený prírastok na 1 000 obyvateľov Natural increase per 1,000 population



Prepočet údajov na územné jednotky platné v roku 2001.
Data recalculated to territorial units valid in 2001.

Prirodzený prírastok Natural increase



Nejvyšší zjištěná hodnota (v kartodiagramu)

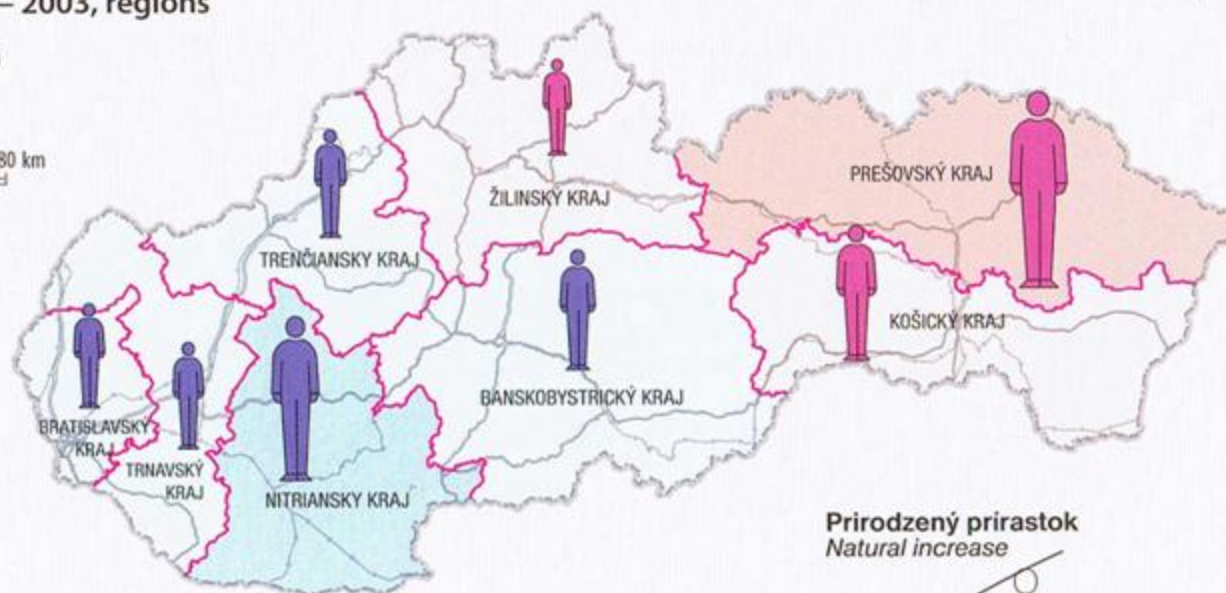
Prirodzený prírastok obyvateľstva priemer za roky 2001 – 2003, kraje

Natural population increase,
average of 2001 – 2003, regions

Jana Marenčáková

1 : 4 000 000

0 40 80 km



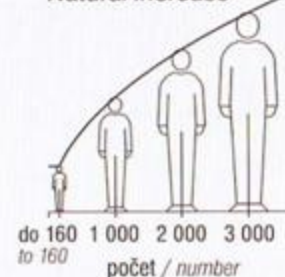
Prírastok na 1 000 obyvateľov
Natural increase per 1,000 population

-2,9 - (-2,1)	0,0 - 1,9
-2,0 - (-0,1)	2,0 - 4,9

prírastok
natural decrease

prírastok
natural increase

Prírastok
Natural increase



Nejvyšší zjištěná hodnota (Výška? Barva?)

Populační povrchy v letech 1921, 1961, 1980 a 2001

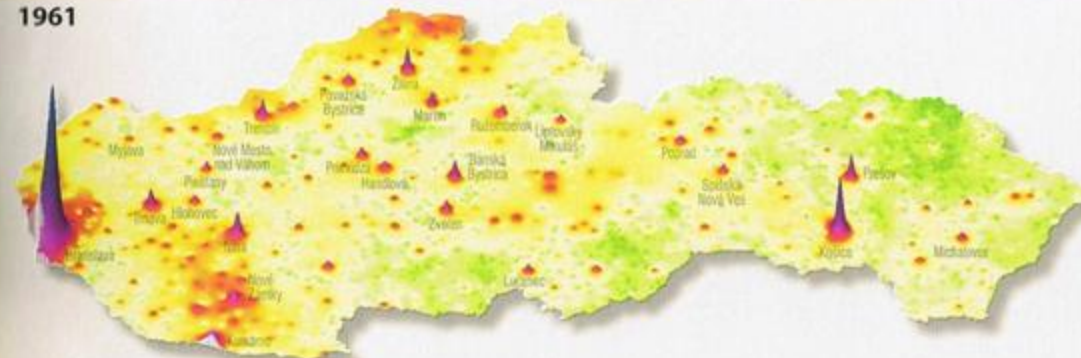
Population surfaces, 1921, 1961, 1980, and 2001

Dagmar Kusendová, Miloslav Ďurica

1921



1961



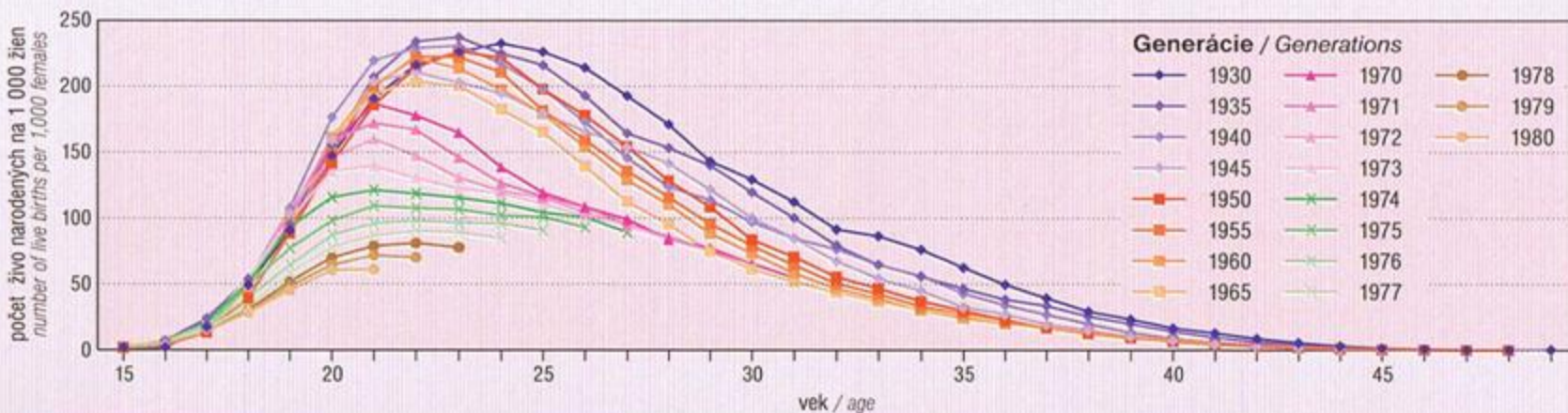
Počet obyvatelů (v tis.)
Population (in thous.)

Extrémní hodnoty—blízko teoretickému rozdělení

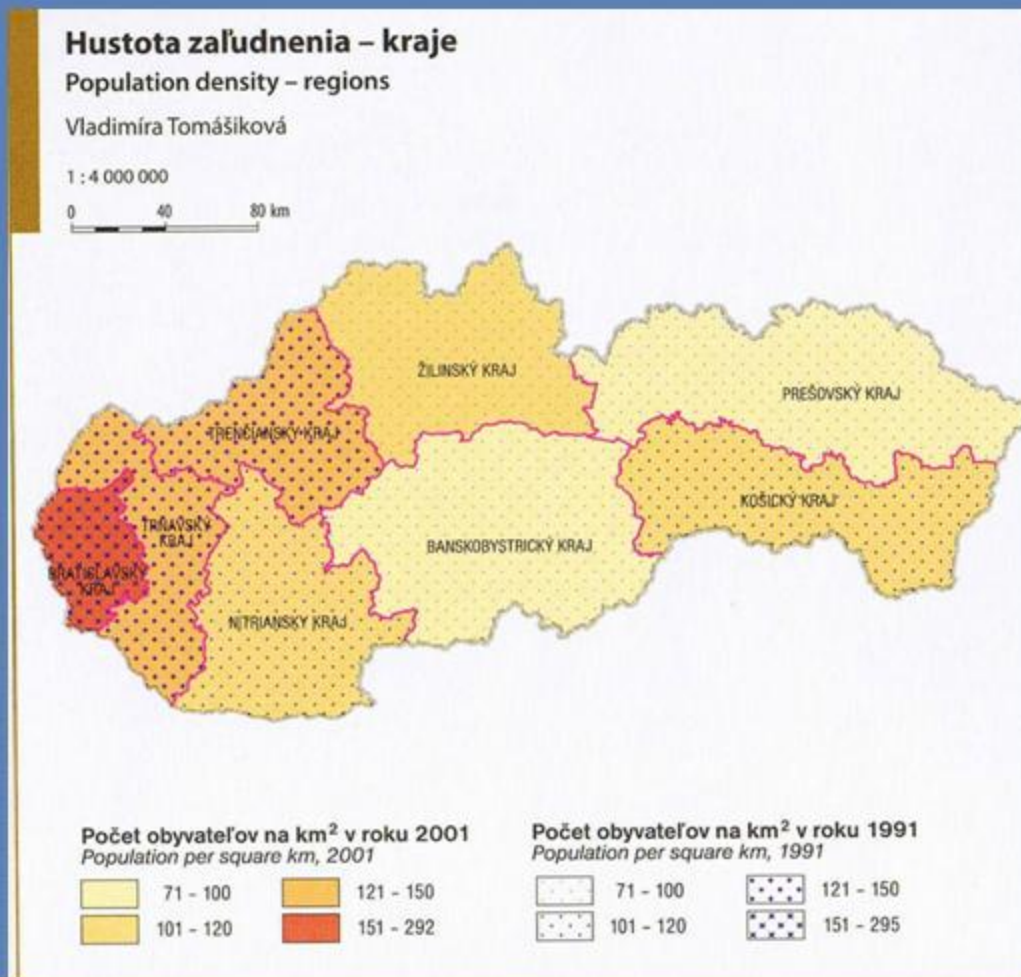
Generační plodnost podle věku matky

Cohorte fertility by age of mother

Jana Marenčáková



Extrémní hodnotu nelze zjistit



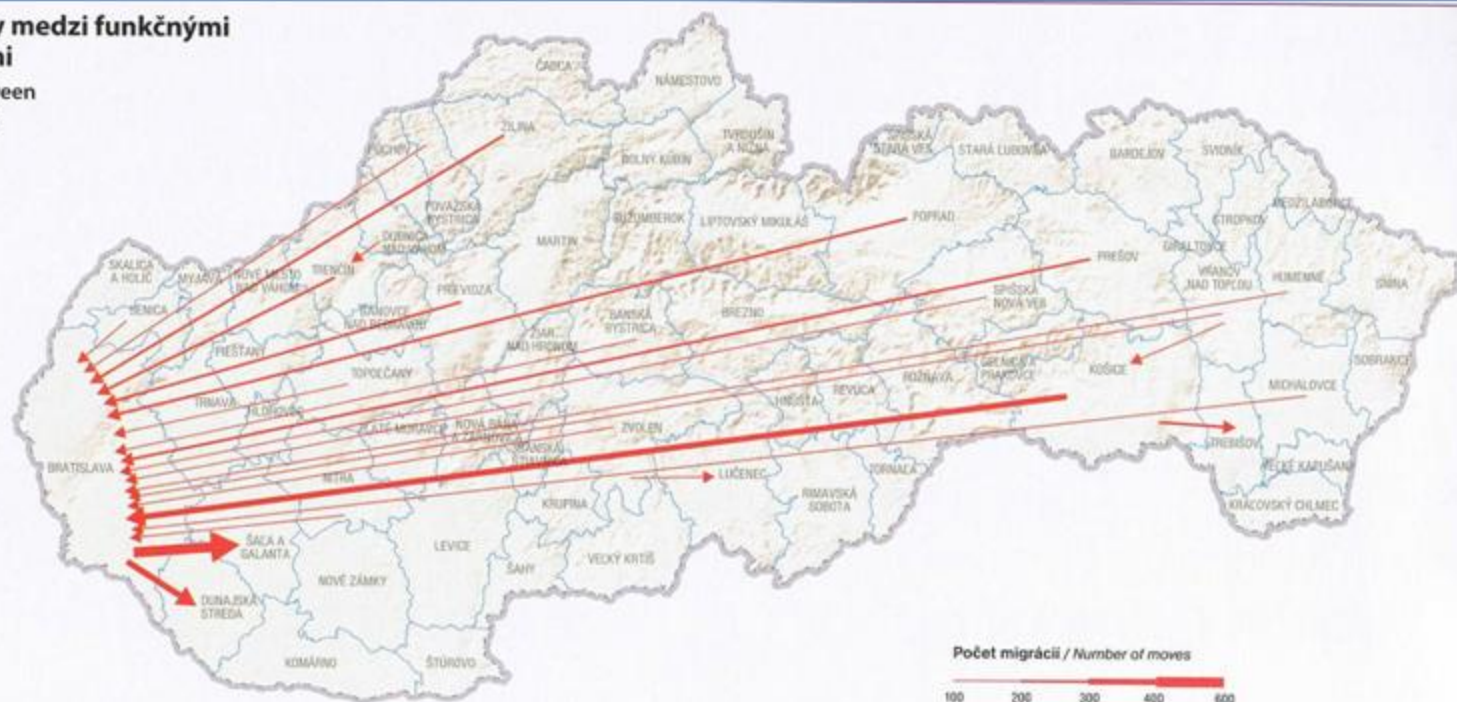
Extrémní hodnotu nelze zjistit

Čisté migrační toky mezi funkčními mestskými regióny

Net migration flows between
functional urban regions

1 : 2 000 000

0 10 20 30 40 50 km



Extrémní hodnotu nelze zjistit

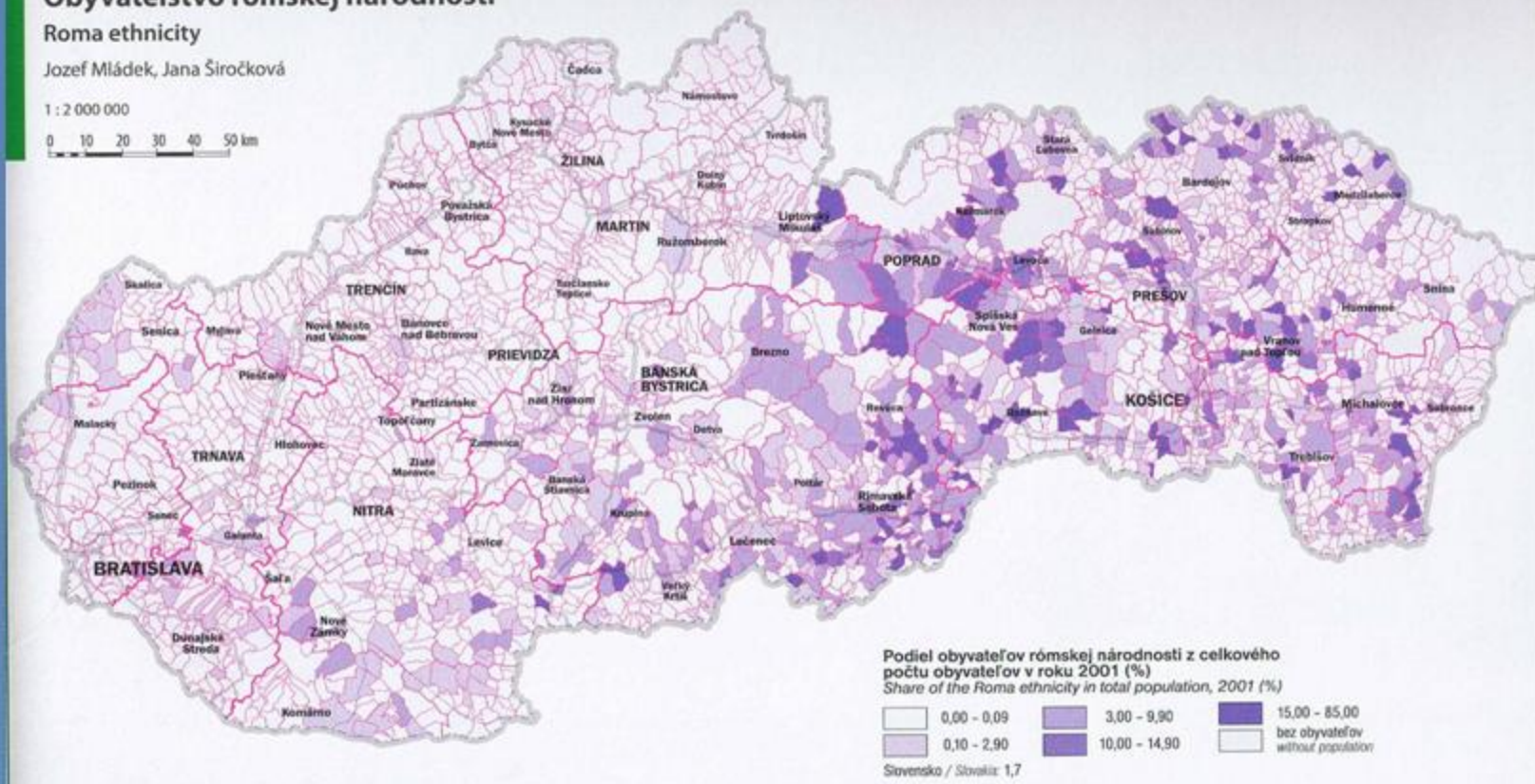
Obyvatelstvo rómskej národnosti

Roma ethnicity

Jozef Mládek, Jana Širočková

1 : 2 000 000

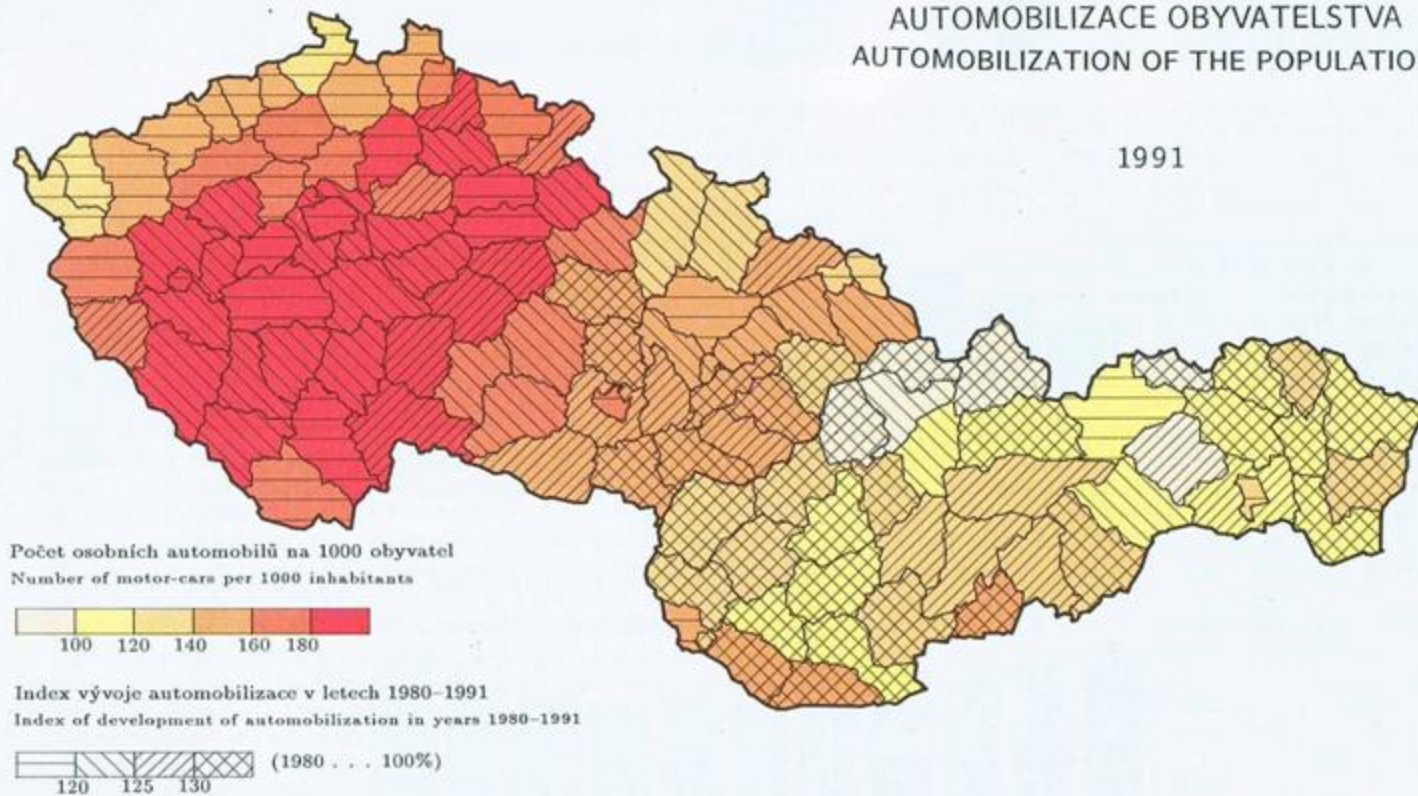
0 10 20 30 40 50 km



Správně řešená stupnice provede regionalizaci

AUTOMOBILIZACE OBYVATELSTVA AUTOMOBILIZATION OF THE POPULATION

1991






DOPORUČENÍ:

Rozlišovat termíny:

- frekvenční extrém
 - datový extrém
 - nejnižší naměřená (zjištěna) hodnota
 - nejvyšší naměřená (zjištěna) hodnota
- 




DOPORUČENÍ:

- V kartogramech a kartodiagramech, (v krajních intervalech), uvádět buď:
 - konkrétní naměřená maxima a minima,
 - nebo teoretické extrémy (pokud existují).
 - A to graficky a hodnotou.
 - Uvádět jak byla stupnice vytvořena.
- 



ZÁVĚRY:

- Nejvyšší a nejnižší naměřená hodnota nemusí být extrém.
 - Kartogram ani kartodiagram neukazují, „samy o sobě“, extrémy.
 - Pro extrémy musíme udělat zvláštní kartografické vyjádření.
 - Nutno odlišit outlier.
 - U každé stupnice uvádět, jakým způsobem byla vytvořena
- 

PODĚKOVÁNÍ

Vít Voženílek

vit.vozenilek@upol.cz

585634513

Jaromír Kaňok

jaromir.kanok@upol.cz

585634519

Pavel Tuček

pavel.tucek@upol.cz

585634521

Děkujeme Vám za pozornost

