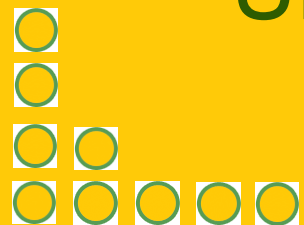




Ukládání dat biodiverzity a jejich vizualizace

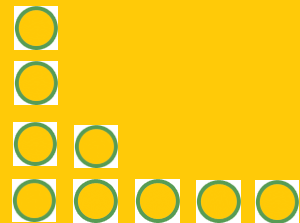
Jiří Jarkovský, Danka Haruštiaková

Institut biostatistiky a analýz, Masarykova univerzita, 2014





II-S1. Ukládání dat



DATA – ukázka uspořádání datového souboru

Parametry (znaky)



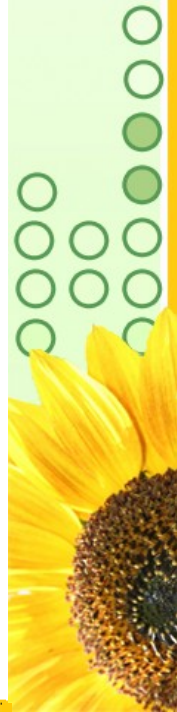
Opakování



Pačet	Uvek	a1u cel.10%	aly% %	a2% %	a3u% %	aly% %	aly cel.10%	a2 cel.10%	a3u cel.10%	aly cel.10%	a4c %	a1sk m/s10 ³	a1Nus m/s10 ³	a1OZ m/s10 ³	a1NaO m/s10 ³
3	1	4									33	12		32	
4	2	7,6	8	58	66	24	06	44	50	1,8	33	95	19	48	10
8	3	4	3	52	55	40	01	21	22	1,6	22	77	35	33	15
11	4	6,1	5	59	64	35	03	36	39	2,1	33	103	26	49	13
12	5	6,9	3	85	88	9	02	59	61	0,6	37	81	13	45	7
14	6	5,9	15	55	70	19	09	33	41	1,1	32	137	33	61	15
16	7	8	18	75	93	7	14	60	74	0,6	34	151	20	59	8
20	8	9,6	3	72	75	23	03	69	72	2,2	40	77	11	38	5
21	9	6	10	67	77	19	06	40	46	1,1	32	120	26	52	11
22	10	3,3	4	55	59	39	01	18	20	1,3	28	81	42	24	12
37	11	3,8	10	60	70	30	04	23	27	1,1	32	111	42	29	11
38	12	6,4	2	76	78	17	01	49	50	1,1	25	356	73	115	23
39	13	6,8	1	57	58	39	01	39	39	2,7	20	234	59	71	18
49	14	8,5	7	67	74	26	06	57	63	2,2	30	156	25	108	17
51	15	9,3	7	57	64	35	07	53	60	3,3	35	129	21	23	4
52	16	2,2	10	55	66	34	02	12	15	0,7	33	46	30	12	8
55	17	9,9	3	78	81	10	03	77	80	0,1	30	189	24	140	18
56	18	5	2	80	82	13	01	40	41	0,7	26	101	25	54	13
6	1	8,8	11	72	83	12	10	63	73	1,1	44	268	366	145	199
9	2	9,2	2	66	68	28	02	61	63	2,6	42	168	269	76	122
13	3	10,0	7	83	90	8	07	83	90	0,8	54	181	201	81	9
15	4	9,6	1	75	76	23	01	72	73	2,2	45	343	47	124	169
17	5	6,0									45	40		21	
19	6	7,2	2	78	80	18	01	56	58	1,3	44	103	178	63	109
24	7	8,2	1	72	73	25	01	59	60	2,1	41	209	349	57	96
26	8	10,3	1	85	86	3	01	88	89	0,3	41	364	41,1	112	126
29	9	5,0	1	74	75	21	01	37	38	1,1	39	83	22,1	32	85
30	10	11,9	1	51	52	47	01	61	62	5,6	33	83	134	52	84
31	11	7,2	3	53	55	29	02	38	40	2,1	28	109	27,1	63	155
32	12	10,8	36	50	76	8	39	54	93	0,9	27	146	15,7	106	114
33	13	11,8	22	54	76	16	26	64	90	1,9	45	246	27,4	63	7
34	14	17,0	1	82	83	16	02	139	141	2,7	34	440	31,2	119	84
40	15	10,0	8	72	80	4	08	72	80	0,4	37	176	220	52	65




Zásady pro ukládání dat

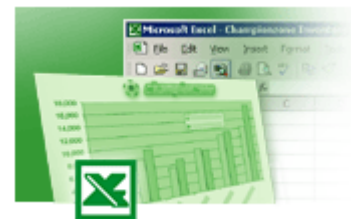
- Správné a přehledné uložení dat je základem jejich pozdější analýzy
- Je vhodné rozmyslet si předem jak budou data ukládána
- Pro počítačové zpracování dat je nezbytné ukládat data v tabulární formě
- Nejvhodnějším způsobem je uložení dat ve formě databázové tabulky
 - Každý sloupec obsahuje pouze jediný typ dat, identifikovaný hlavičkou sloupce
 - Každý řádek obsahuje minimální jednotku dat (např. pacient, jedna návštěva pacienta apod.)
 - Je nepřípustné kombinovat v jednom sloupci číselné a textové hodnoty
 - Komentáře jsou uloženy v samostatných sloupcích
 - U textových dat nezbytné kontrolovat překlepy v názvech kategorií
 - Specifickým typem dat jsou datумы u nichž je nezbytné kontrolovat, zda jsou datумы uloženy v korektním formátu
- Takto uspořádaná data je v tabulkových nebo databázových programech možné převést na libovolnou výstupní tabulku
- Pro základní uložení a čištění dat menšího rozsahu je možné využít aplikaci MS Office







Ukládání dat v MS Office

• MS Excel

-  Kontingenční tabulky – rychlá sumarizace rozsáhlých tabulek
-  Možnost výpočtů a grafových výstupů přímo v aplikaci
-  Visual Basic – složitější aplikace
- Omezení tabulky na 256x65536 buněk (do verze 2003)
- Omezená kontrola chyb při zadávání



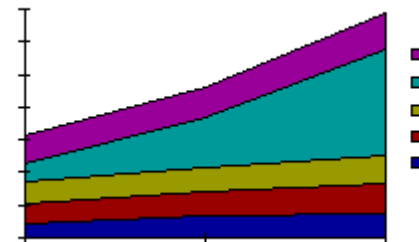
• MS Access

-  Plnohodnotná databáze vhodná pro velké množství dat, řádky omezeny v podstatě jen dostupnou pamětí
-  Kontrola typu dat
-  Relace tabulek – omezení velikosti souboru
-  Visual Basic a formuláře – složitější aplikace
- Omezení tabulky na 255 sloupců
- Výpočty a grafy jsou složitější než v Excelu



Možnosti MS Excel

- Správa a práce s tabulárními daty
- Řazení dat, výběry z dat, přehledy dat
- Formátování a přehledné zobrazení dat
- Zobrazení dat ve formě grafů
- Různé druhy výpočtů pomocí zabudovaných funkcí
- Tvorba tiskových sestav
- Makra – zautomatizování častých činností
- Tvorba aplikací (Visual Basic for Applications)



Počet z	Délka	Pohlaví
1	2	
2		
3		
4		
5		
6		
7	26	
8	106	
9	121	
10	160	
11	34	
12	45	
13	70	
14	72	
15	87	
16	Celkový součet	
17		

17	10	2
18	12	3
19	5	4
20	8	5
21	4	8
22	7	9
23	9	11
24	suma součinů řádků	310
25		



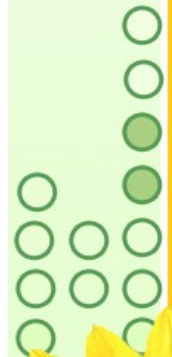
Import a export dat

- **Import dat**

- Manuální zadávání
- import – podpora importu ze starších verzí Excelu, textových souborů, databází apod.
- kopírování přes schránku Windows – vkládání z nejrůznějších aplikací – MS Office, Statistica atd.
- využití textových souborů jako kompatibilního formátu pro přenos dat mezi různými aplikacemi

- **Export dat**

- Ukládáním souborů ve formátech podporovaných jinými SW, časté jsou textové soubory, dbf soubory nebo starší verze Excelu
- Přímé kopírování přes schránku Windows



Tipy a triky

• Výběr buněk

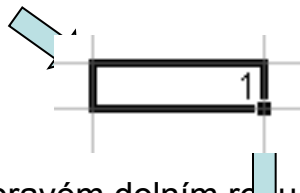
- CTRL+A – výběr celého listu
- CTRL + klepnutí myší do buňky – výběr jednotlivých buněk
- SHIFT + klepnutí myší na jinou buňku – výběr bloku buněk
- SHIFT + šipky – výběr sousedních buněk ve směru šipky
- SHIFT+CTRL+END (HOME) – výběr do konce (začátku) oblasti dat v listu
- SHIFT+CTRL+šipky – výběr souvislého řádku nebo sloupce buněk
- SHIFT + klepnutí na objekty – výběr více objektů

• Kopírování a vkládání

- CTRL+C – zkopírování označené oblasti buněk
- CTRL+V – vložení obsahu schránky – oblast buněk, objekt, data z jiné aplikace

• Myš a okraje buňky

- Chycení myší za okraj umožňuje přesun buňky nebo bloku buněk

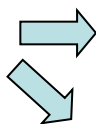


- Při chycení čtverečku v pravém dolním rohu výběru je tažením možno vyplnit více buněk hodnotami původní buňky (ve vzorcích se mění relativní odkazy, je také možné vyplnění hodnotami ze seznamu – např. po sobě jsou názvy měsíců).

Databázová struktura dat v Excelu

Sloupce tabulky = parametry záznamů, hlavička udává obsah sloupce – stejný údaj v celém sloupci

Jednotlivé záznamy
(taxon, lokalita atd.)

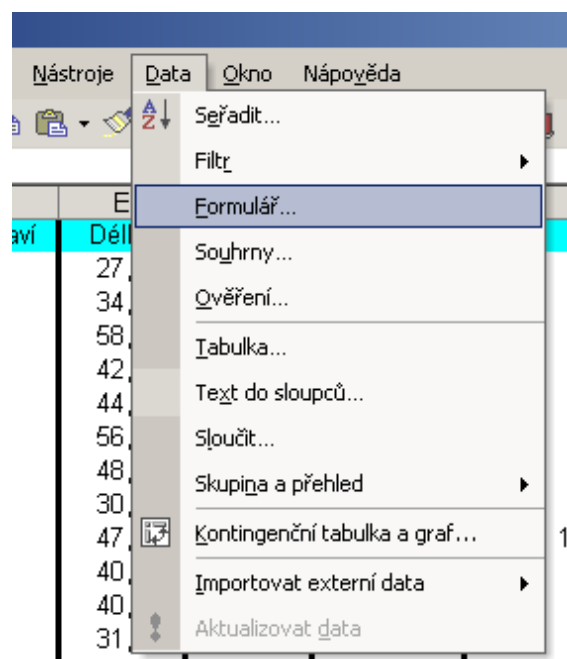


	A	B	C	D	E	F	G	H	I
1	Číslo	Značka	Společ	Pohlaví	Délka	Váha	P. anguillae	P. bini	
2	1	1	1	m	27,5	23,0	2	2	
3	2	2	2	f	34,0	62,5	0	2	
4	3	5	3	f	58,0	230,0	0	0	
5	4	6	4	f	42,0	155,0	0	0	
6	5	7	5	f	44,0	149,8	0	0	
7	6	8	6	f	56,0	323,0	0	1	
8	7	9	7	m	48,5	178,2	0	0	
9	8	10	8	f	30,5	47,7	4	6	
10	9	11	9	f	47,0	175,9	5	14	
11	10	12	10	f	40,0	85,1	5	9	
12	11	14	11	f	40,0	101,0	0	0	
13	12	15	12	f	31,0	84,0	15	9	
14	13	16	13	f?	22,0	9,0	0	0	
15	14	17	14	f	42,0	108,0	1	3	
16	15	18	15	f	44,0	130,0	0	0	
17	16	19	16	f	37,0	85,0	2	5	
18	17	20	17	f	50,0	212,0	1	8	



Automatický zadávací formulář

- Slouží k usnadnění zadávání dat do databázových tabulek
- Načítá automaticky hlavičky sloupců jako zadávané položky



The 'data' dialog box contains the following fields and controls:

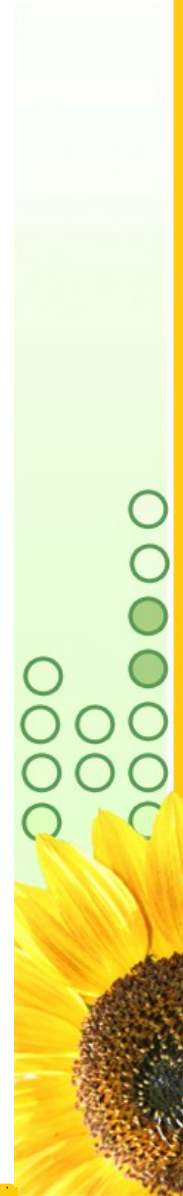
- Číslo ryby: 1 (with a scroll bar and '1 z 19' indicator)
- Značka ryby: 1
- Společ číslo: 1
- Pohlaví: m
- Délka: 27,5
- Váha: 23
- P. anguillae: 2
- P. bini: 2
- Buttons: Nový, Odstranit, Obnovit, Předchozí, Další, Kritéria, Zavřít

Nový záznam

Vyhledávání

Názvy sloupců

Obsah dané buňky - editovatelný



Automatické seznamy

- Vytváří se z hodnot buněk v daném sloupci a umožňují vložit hodnotu výběrem ze seznamu již zadaných hodnot – usnadnění zadávání

Sloupec z něj je seznam vytvořen a pro který platí

lacin	urbane	Locata	ec

Buňka, do níž se vloží vybraná hodnota

Glo

- Vyjmout
- Kopírovat
- Vložit
- Vložit jinak...
- Vložit buňky...
- Odstranit...
- Vymazat obsah
- Vložit komentář
- Formát buněk...
- Vybrat ze seznamu...
- Přidat kukátko
- Hypertextový odkaz...

Caryophyllaeides fennica (Schneider, 1902)

Piscicola geometra (Linnaeus, 1761)

Acanthocephallus lucii (Müller, 1776)

Apophallus mühlungi Jägerskiöld, 1899

Argulus foliaceus (Linnaeus, 1758)

Caryophyllaeides fennica (Schneider, 1902)

D. cabaleroi

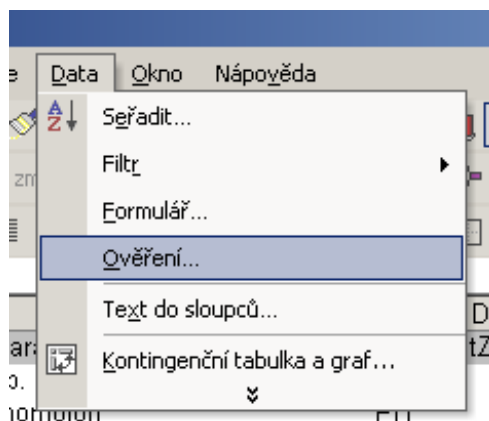
D. crucifer Wagener, 1857

D. fallax Wagener, 1857

D. nanus Dogiel et Bychowsky, 1934

Automatická kontrola dat

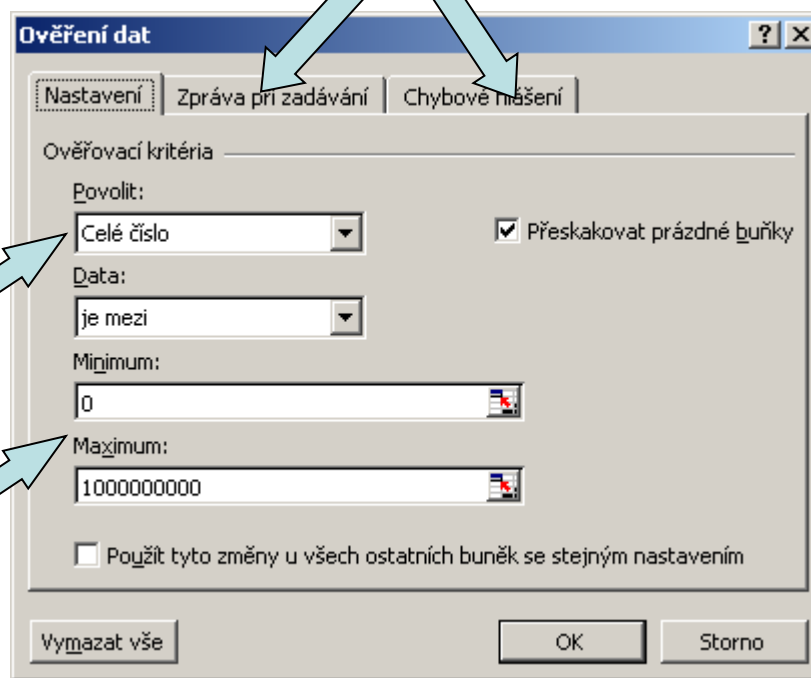
- Umožňuje ověřit typ, rozsah nebo povolit pouze určitý seznam hodnot zadávaných do sloupce databázové tabulky



Co je povoleno – definiční obory čísel, seznamy, vzorce atd.

Rozsahy hodnot, načtení seznamů apod.

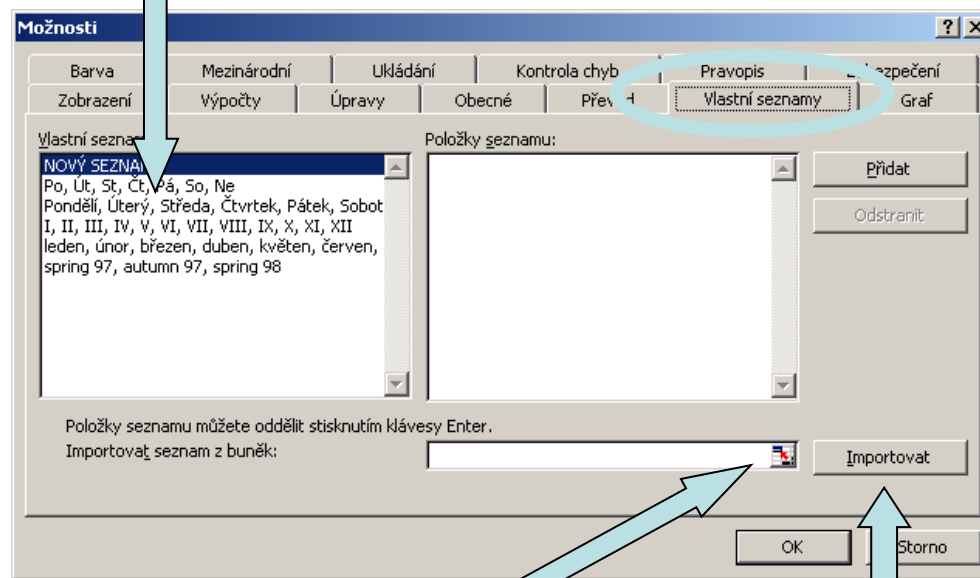
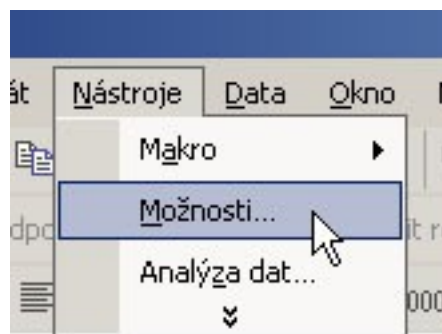
komunikace s uživatelem



Seznamy

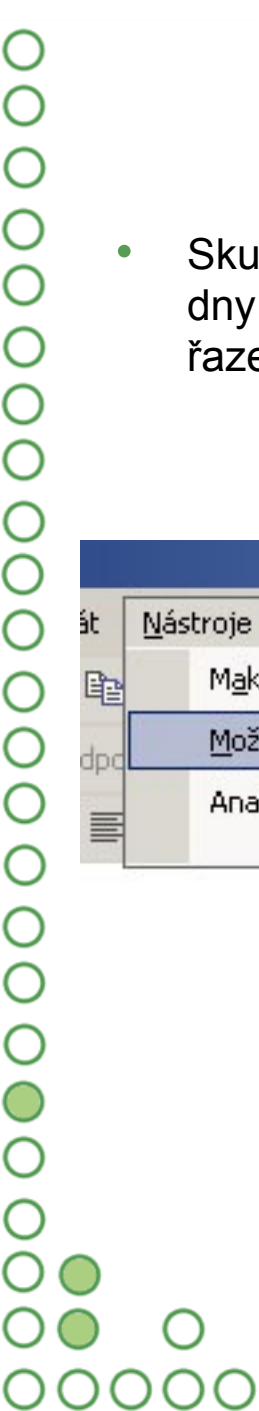
- Skupiny hodnot zachovávající logické pořadí, některé jsou zabudované (např. dny v týdnu, měsíce v roce), další je možné uživatelsky vytvořit, slouží pro účely řazení a automatického vyplňování dat

Existující seznamy



Výběr buněk pro nový seznam

Načtení nového seznamu



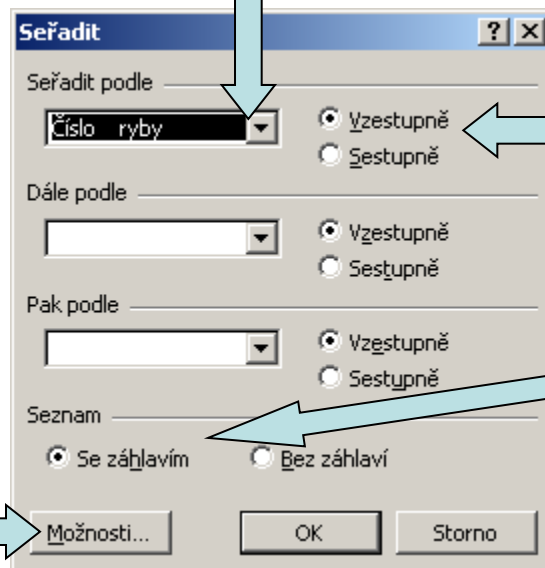
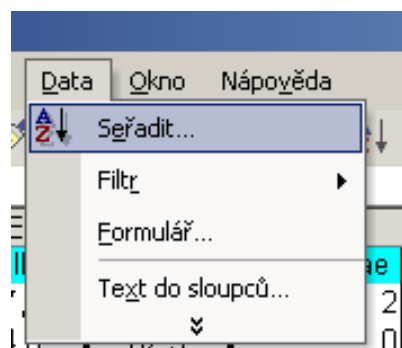
Řazení dat

- Řazení dat je nejjednodušším způsobem jejich zpřehlednění, užitečným hlavně u menších/výsledkových tabulek

Zkontrolujte, zda seřazení nezničí vazby mezi buňkami = kontrola oblasti, kterou řadíte.



Podle čeho řadit



Směr řazení – vzestupně, sestupně

Využít první řádek oblasti jako záhlaví

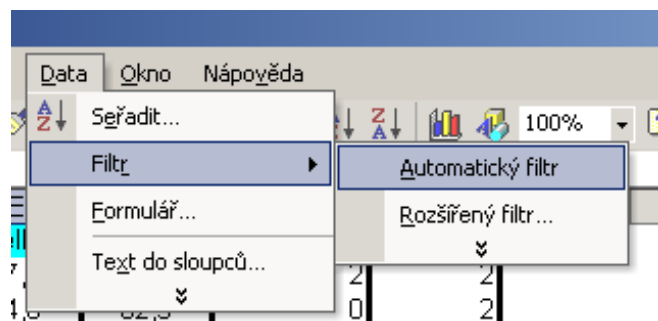
Další možnosti – řazení řádků, řazení podle seznamu



Automatický filtr

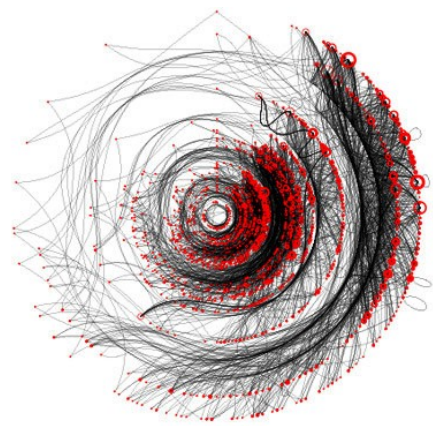
- Pomocí automatického filtru je snadné vybírat úseky dat pro další zpracování na základě hodnot ve sloupcích databázové tabulky, výběr je možný i podle více sloupců (např. určitá skupina pacientů)
- Funkce automaticky rozezná hlavičky sloupců v souvislé oblasti buněk
- U sloupců použitých pro filtraci jsou rozbalovací seznamy zbarveny modře
- **Výhodné pro čištění dat (vyhledávání překlepů, kombinace textu a čísel)**

Výběr hodnot pro filtraci

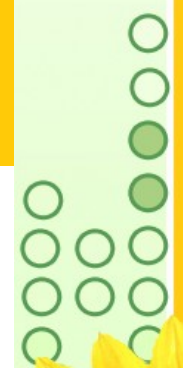
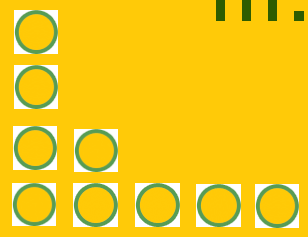


Rozbalení seznamu hodnot nalezených ve sloupci

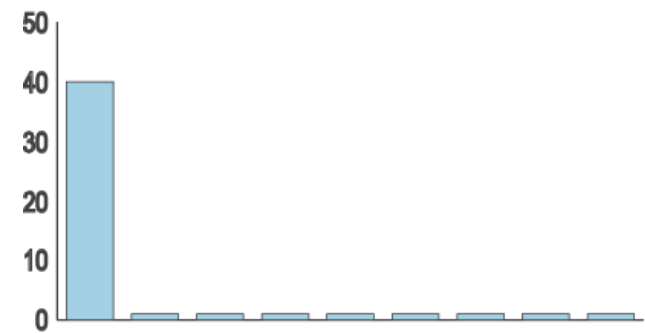
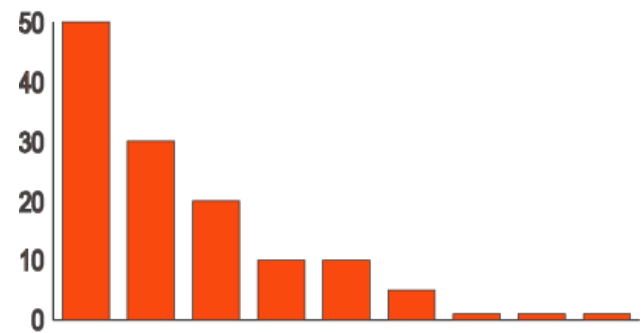
	A	B	C	D	E
1	Číslo	Značka	Společ	Pohlaví	Délka
2	1	1	1	(Vše)	27,5
3	2	2	2	(Prvních 10...)	34,0
4	3	5	3	f	58,0
5	4	6	4	f?	42,0
6	5	7	5	m	44,0
7	6	8	6	f	56,0
8	7	9	7	m	48,5



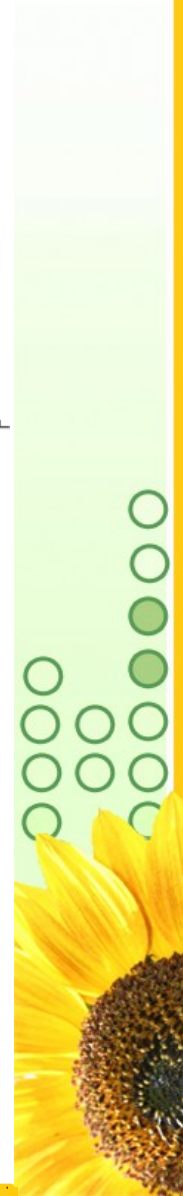
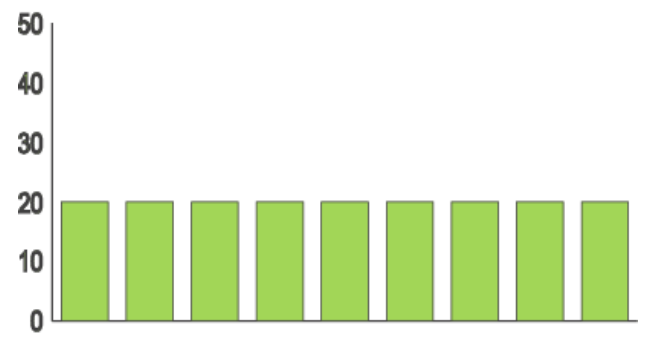
III. Vizualizace biodiverzitních dat



Species abundance curves

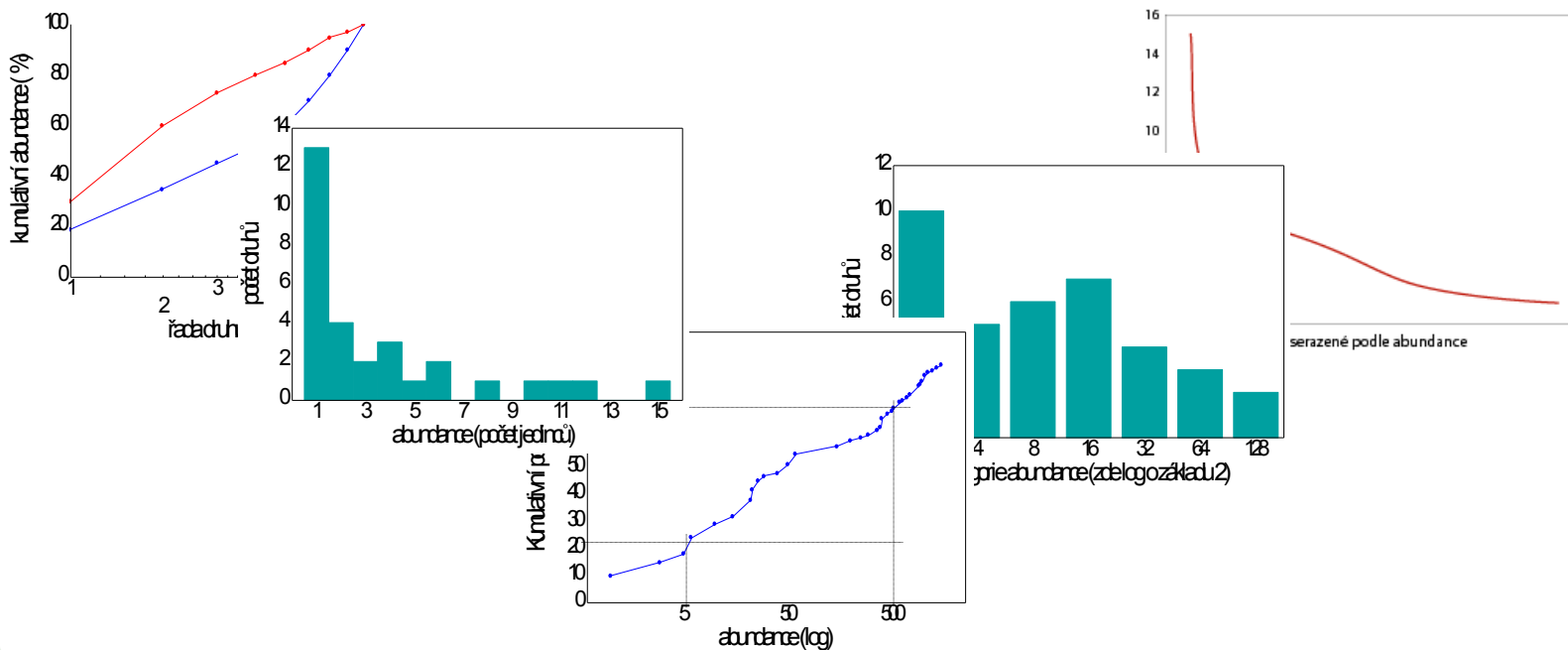


?



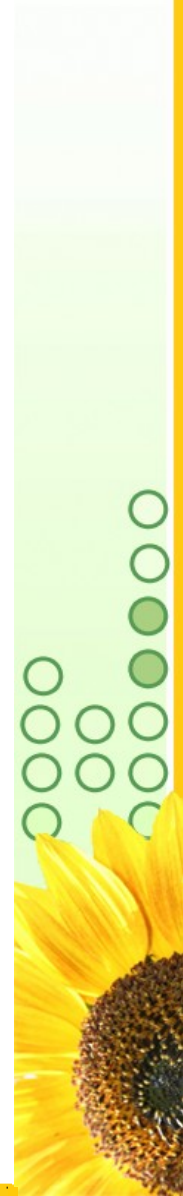
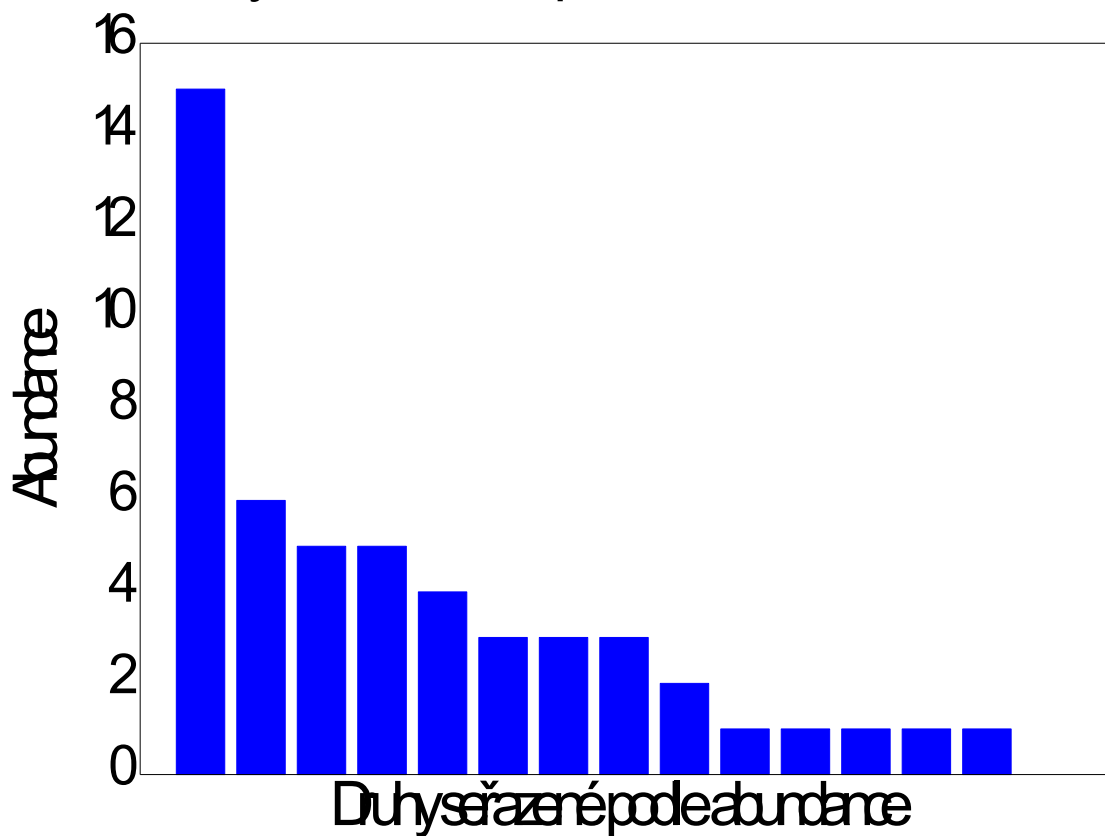
Zobrazení křivky společenstva

- Různé metody zobrazení – různé pohledy na společenstvo
- Jednoduché optické srovnání různých společenstev nebo modelových průběhů společenstev
- Jednoduchá tvorba v Excelu



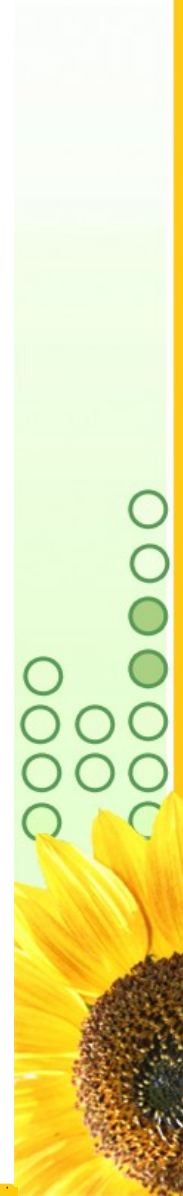
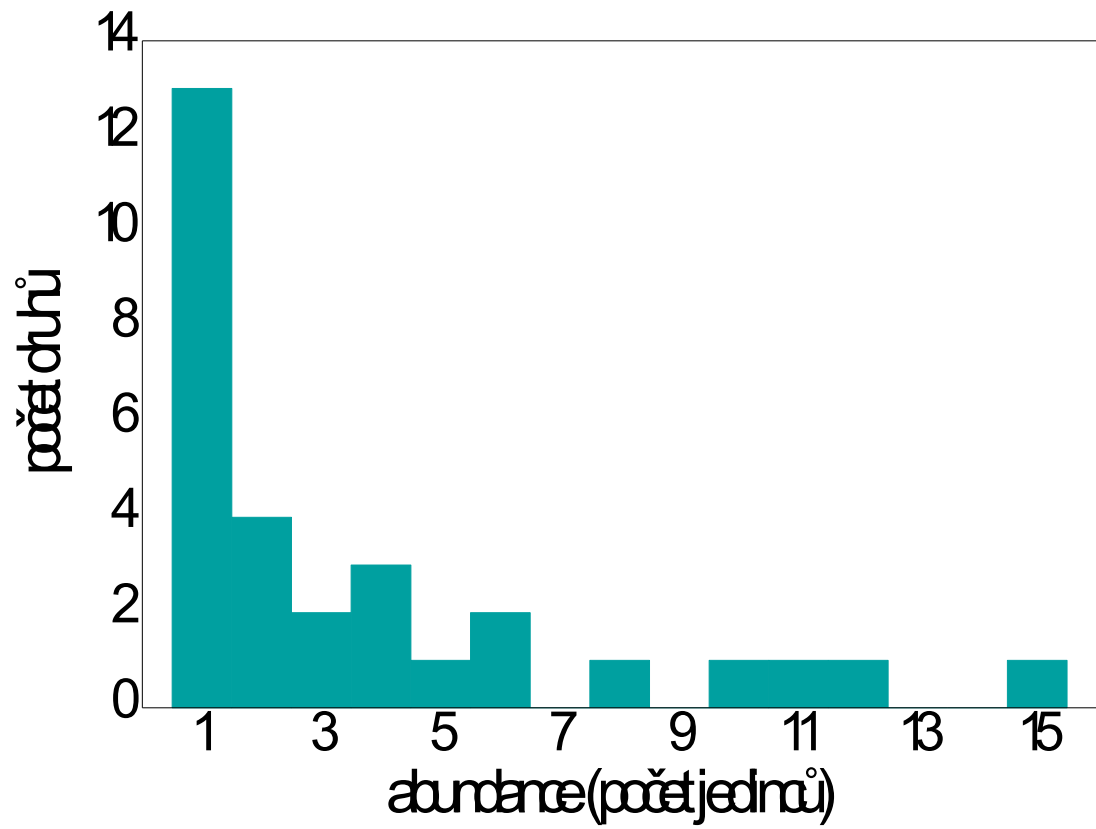
Graf pořadí abundancí

- Graf zobrazuje abundanci druhů seřazených podle této abundance. Poskytuje uživateli přehled o tvaru společenstva – vyrovnanost abundancí, přítomnost „ocásku“ vzácných druhů apod.



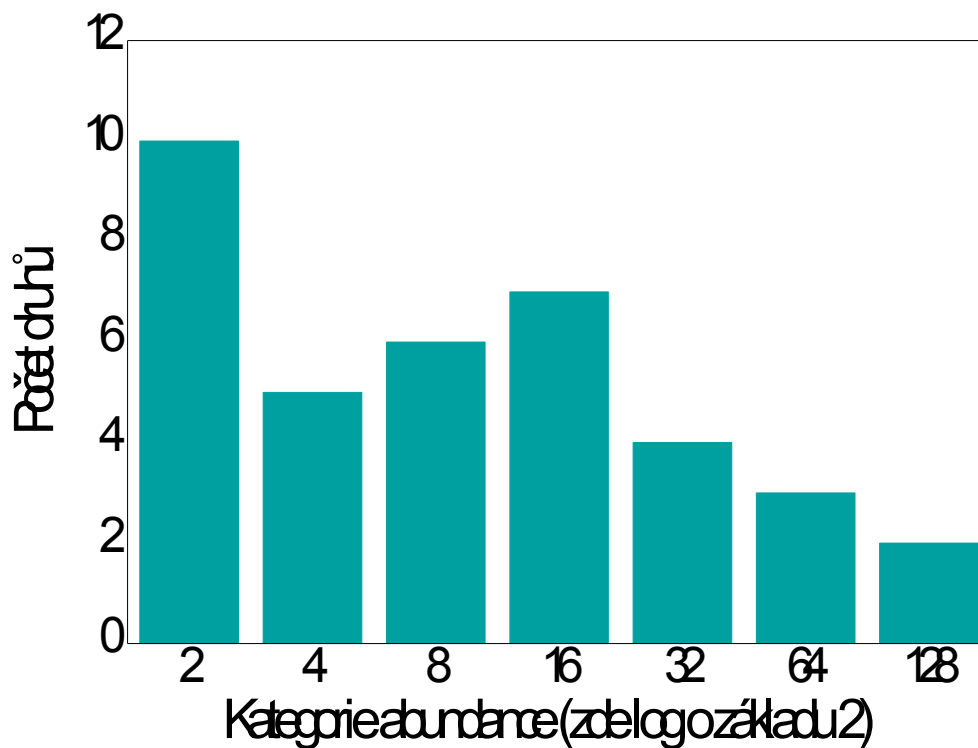
Graf abundance druhů

- Četnost druhů s danou abundancí je vynášena proti hodnotám nalezených abundancí. Graf poskytuje uživateli přehled o rozložení vzácných, středně početných a hojně se vyskytujících druhů.



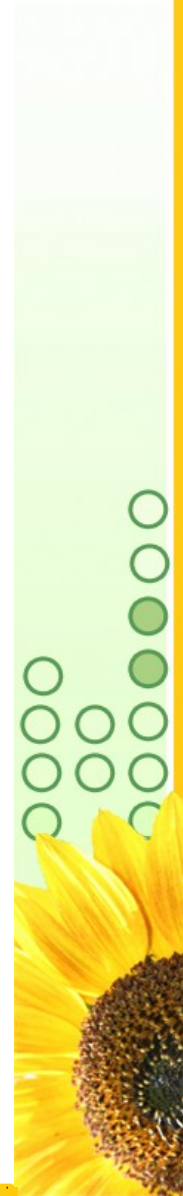
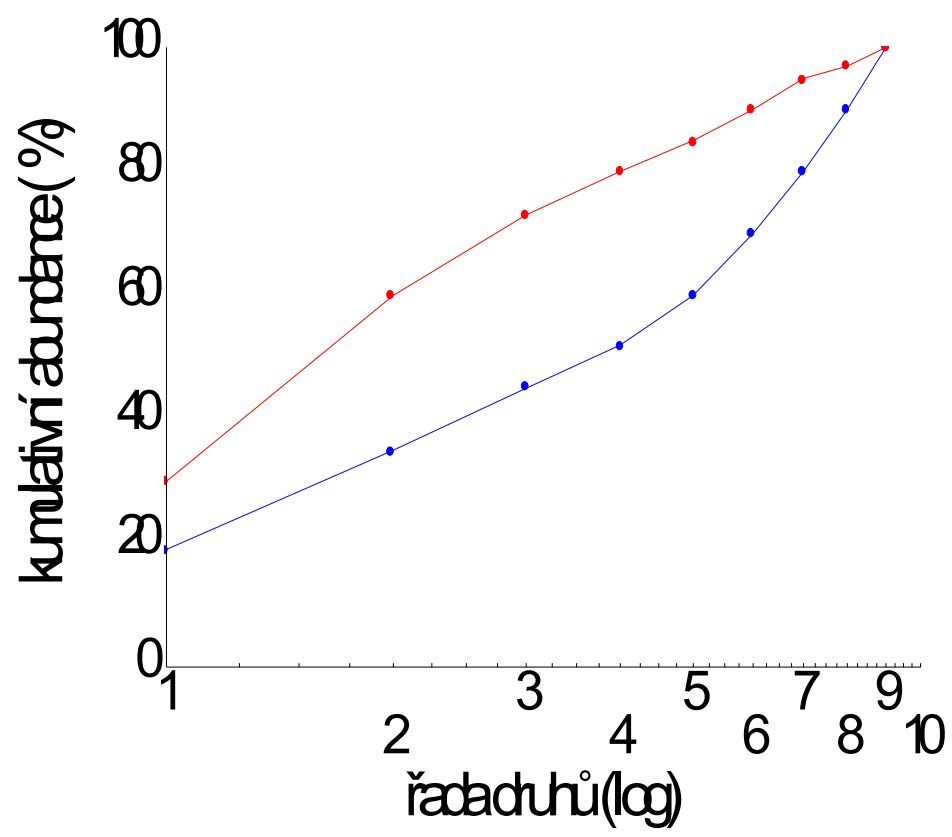
Graf kategorií abundance

- Je obdobou grafu četnosti druhů s určitou abundancí, namísto konkrétních hodnot zde jsou ale třídy abundance a četnost druhů k nim náležejících. Opět umožňuje sledovat relativní podíly vzácných a hojných druhů. Ve formě početnosti druhů v log třídách abundance jsou generovány výsledky některých species abundance models.



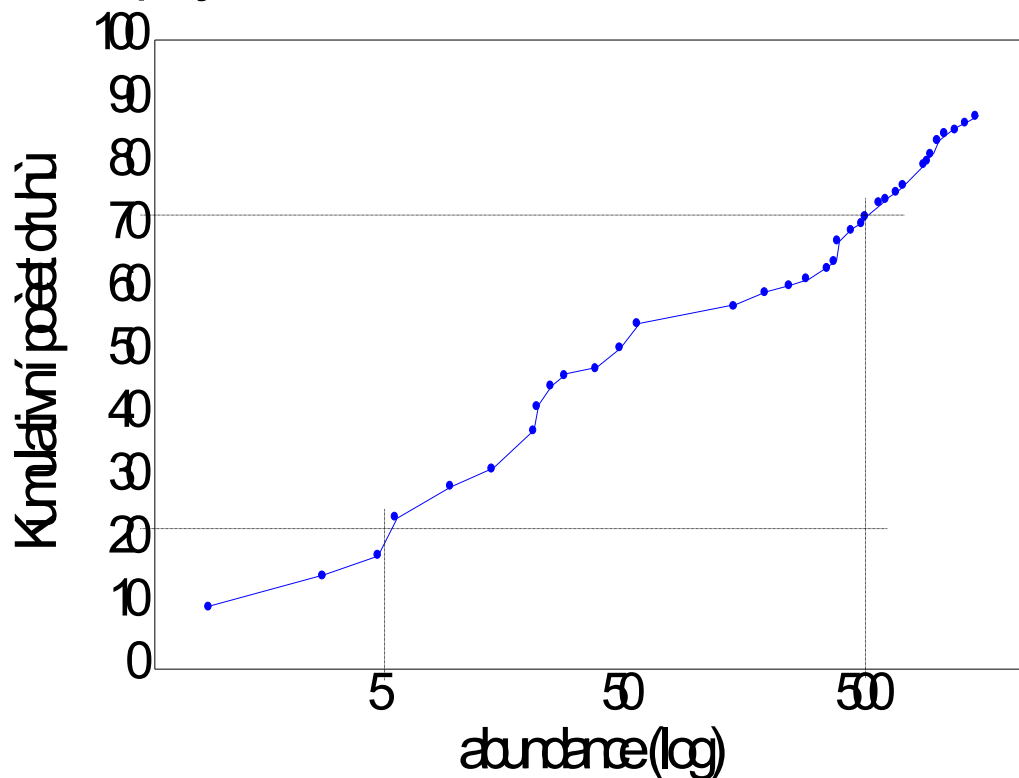
K- dominance plot

- Tento graf vynáší kumulativní abundanci druhů proti logaritmu druhové řady. Může být využit pro optické srovnání diverzity různých vzorků



Kumulativní počet druhů

- Graf vynáší kumulativní počet druhů proti ose jejich logaritmované abundance. Graf slouží jako doplněk k výpočtu Q statistiky, který je založen na obdobně uspořádaných datech. Zobrazuje strmost narůstání počtu druhů se stoupající abundancí.

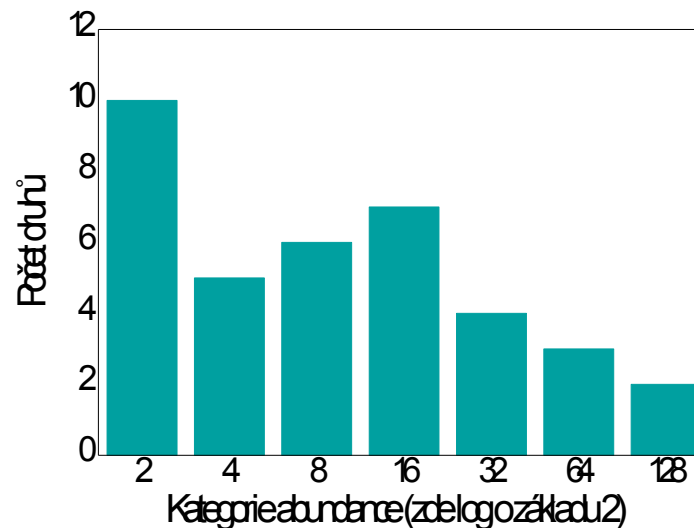


Rozložení dat

- Abundance taxonů napříč lokalitami nabývají obvykle lognormálního rozdělení
- Častý výskyt odlehlých hodnot (lokalita s taxonem s vysokou abundancí)



Pozor na průměr!
Lepší bude medián!



Transformace a standardizace dat

- Logaritmická transformace o vhodném základu (nejčastěji přirozený nebo dekadický)

$$y'_i = \log(b_0 + b_1 y_i) \quad b_0 = 1, b_1 = 1$$

- Při srovnání diverzity mezi různými společenstvy je třeba zohlednit i rozdílnou celkovou abundanci a data před srovnáním standardizovat.
 - Přepočet abundancí na procentuální strukturu společenstva
 - Standardizace abundance každého taxonu na standardní normální rozdělení

$$z_i = \frac{y_i - \bar{y}}{s_y}$$

