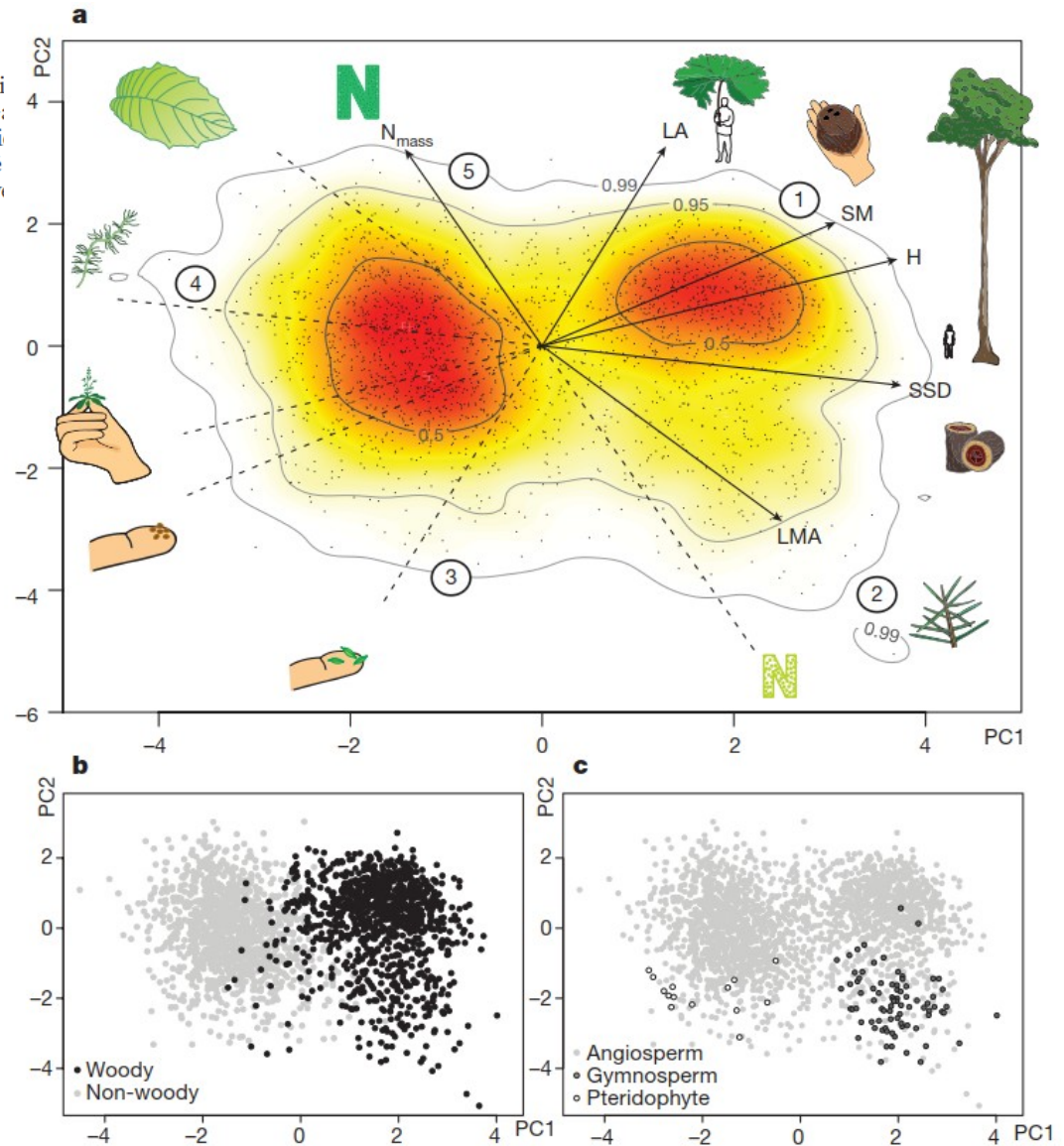


# ZPRACOVÁNÍ DAT V EKOLOGII SPOLEČENSTEV

JAKUB TĚŠITEL S POUŽITÍM MATERIÁLŮ VÍTA SYROVÁTKY A DAVIDA ZELENÉHO

# The global spectrum of plant form and function

Sandra Díaz<sup>1</sup>, Jens Kattge<sup>2,3</sup>, Johannes H. C. Cornelissen<sup>4</sup>, Ian J. Wri Michael Kleyer<sup>10</sup>, Christian Wirth<sup>2,3,11</sup>, I. Colin Prentice<sup>5,12</sup>, Eric G. Hendrik Poorter<sup>14</sup>, Peter B. Reich<sup>15,16</sup>, Angela T. Moles<sup>17</sup>, John Dicki Jérôme Chave<sup>22</sup>, S. Joseph Wright<sup>23</sup>, Serge N. Sheremet'ev<sup>24</sup>, Hervé Simon Pierce<sup>30</sup>, Bill Shipley<sup>31</sup>, Donald Kirkup<sup>32</sup>, Fernando Casanov Nadja Rüger<sup>3,23</sup>, Miguel D. Mahecha<sup>2,3</sup> & Lucas D. Gorné<sup>1</sup>

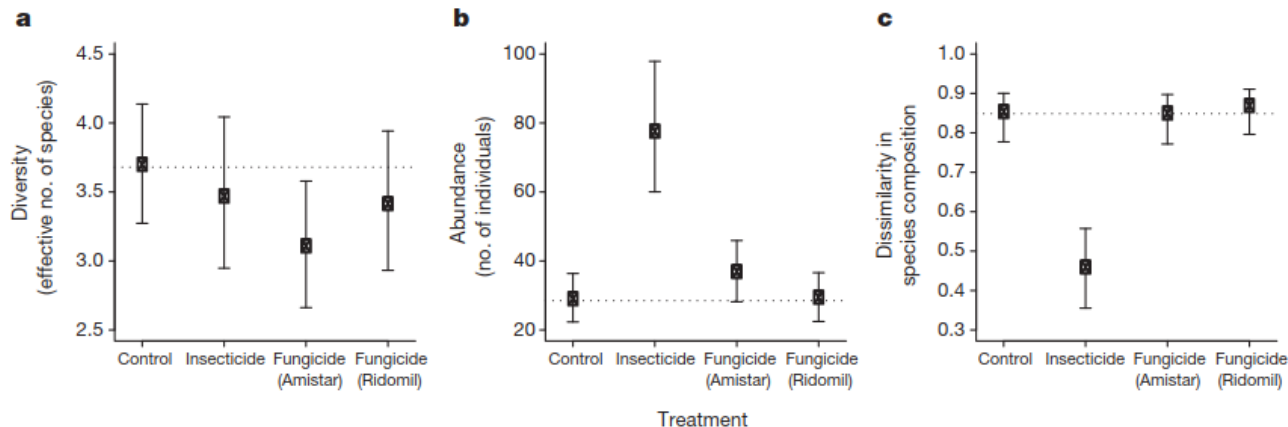


# Pathogens and insect herbivores drive rainforest plant diversity and composition

Robert Bagchi<sup>1,2</sup>, Rachel E. Gallery<sup>1,3</sup>, Sofia Gripenberg<sup>1,4</sup>, Sarah J. Gurr<sup>5,6</sup>, Lakshmi Narayan<sup>1</sup>, Claire E. Addis<sup>1</sup>, Robert P. Freckleton<sup>7</sup> & Owen T. Lewis<sup>1</sup>



## RESEARCH LETTER



**Figure 1 | Suppression of insects and pathogens alters seedling community composition and diversity, respectively.** **a–c**, Effects of insecticide (Engeo) and two fungicides (Amistar and Ridomil) on the mean effective number of species recruiting as seedlings (**a**); the mean seedling abundance summed

across all species (**b**); and the mean abundance-weighted Morisita–Horn dissimilarity in species composition for seedlings under each treatment compared to seeds in adjacent seed traps (**c**). The error bars represent the 95% confidence intervals of the mean across the 36 stations.

# Patterns of bryophyte and vascular plant richness in European subalpine springs

Lucia Sekulová · Michal Hájek · Petra Hájková ·  
Eva Mikulášková · Alexandre Buttler ·  
Vít Syrovátka · Zuzana Rozbrojová

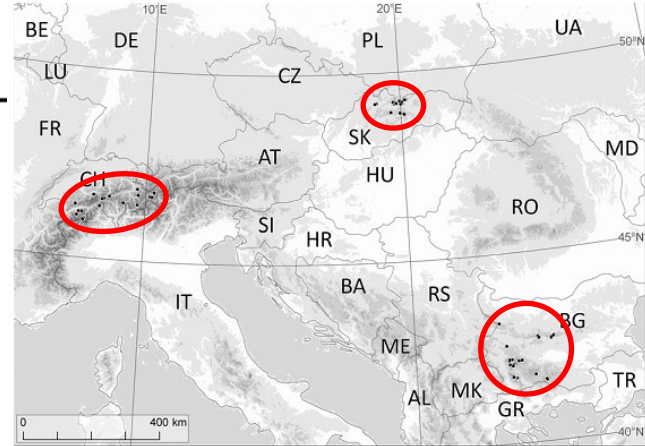
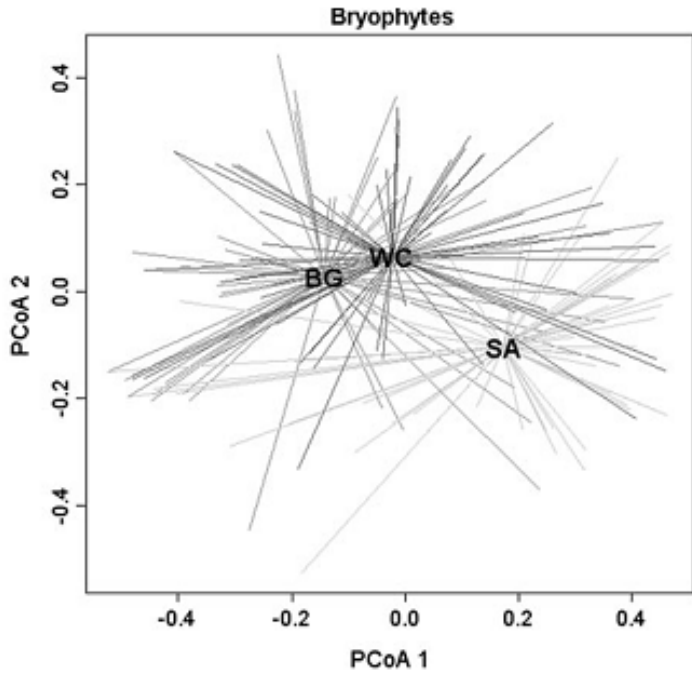
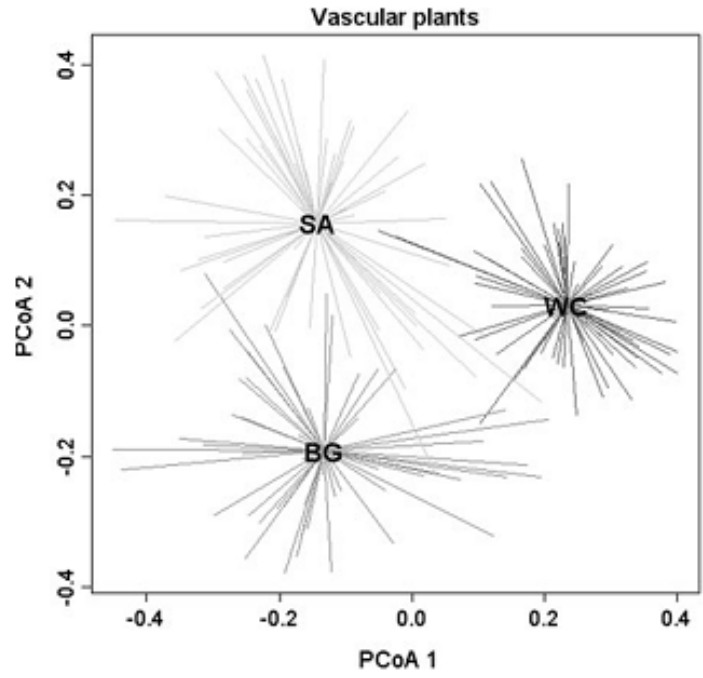




Fig. 1 The location of vegetation plots in the study regions. The country codes are indicated

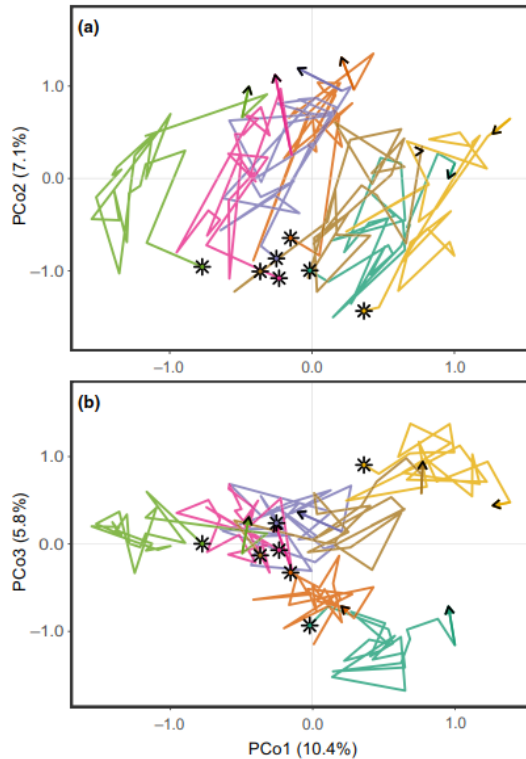


**Fig. 4** Comparison of overall  $\beta$ -diversity of spring vegetation in the three study regions for vascular plants and bryophytes.  $\beta$ -Diversity (dispersion of groups) of vascular plants was highest in Swiss Alps and lowest in Western Carpathians while no

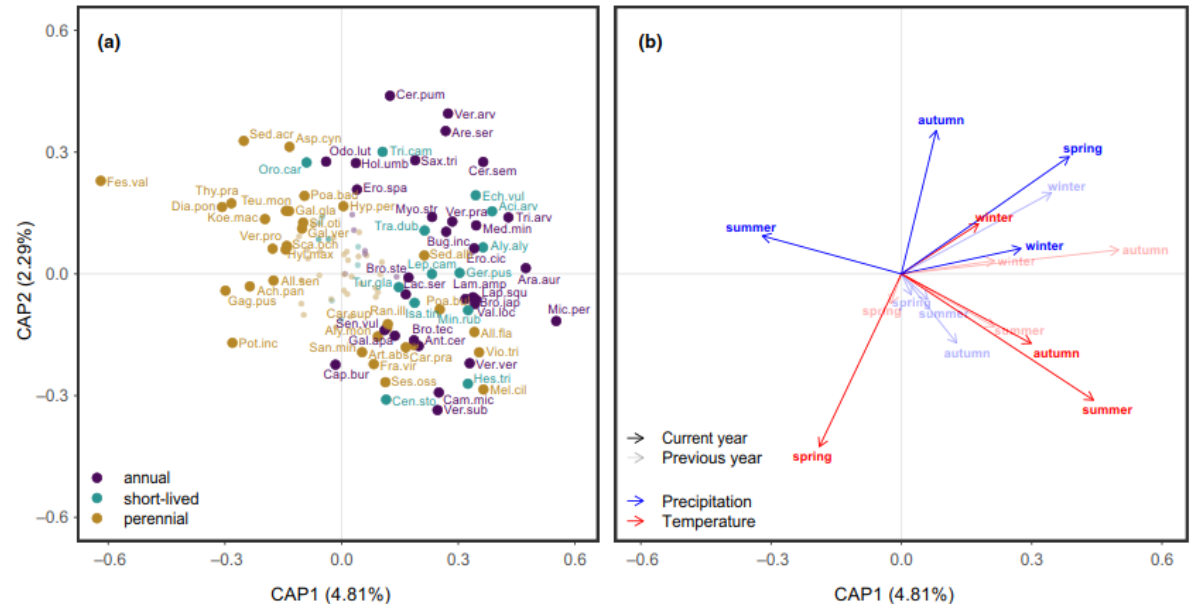
difference was found for bryophytes (ANOVA and subsequent Tukey's post hoc test). *BG* Bulgarian mountains, *SA* Swiss Alps, *WC* Western Carpathians

# Weather fluctuations drive short-term dynamics and long-term stability in plant communities: A 25-year study in a Central European dry grassland

Felícia M. Fischer<sup>1</sup>  | Kryštof Chytrý<sup>1</sup>  | Jakub Těšitel<sup>1</sup>  | Jiří Danihelka<sup>1,2</sup>  | Milan Chytrý<sup>1</sup> 



**FIGURE 5** Ordination (principal coordinate analysis) of repeated surveys of seven permanent plots on Děvín Hill over 25 years. (a) First and second axis, (b) first and third axis. Each plot is indicated by a unique colour with the lines connecting the consecutive years. The asterisks and arrowheads represent the start (1993) and end (2018) years of the repeated survey, respectively [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]



**FIGURE 7** Ordination plots of partial distance-based redundancy analysis (db-RDA), showing the effect of precipitation and temperature on the dynamics of grassland community composition on the Děvín Hill. Plot identities were used as covariates in the model. Effects of precipitation and temperature are indicated by blue and red, respectively. Dark and light colours represent the current-year seasons and previous-year seasons, respectively (compare Figure 2). Results of Monte-Carlo permutation test for all constrained axes:  $p = 0.001$  (under

# Comparison of diatom community structure from epilithon and fish guts: implications for inferring past changes in water quality

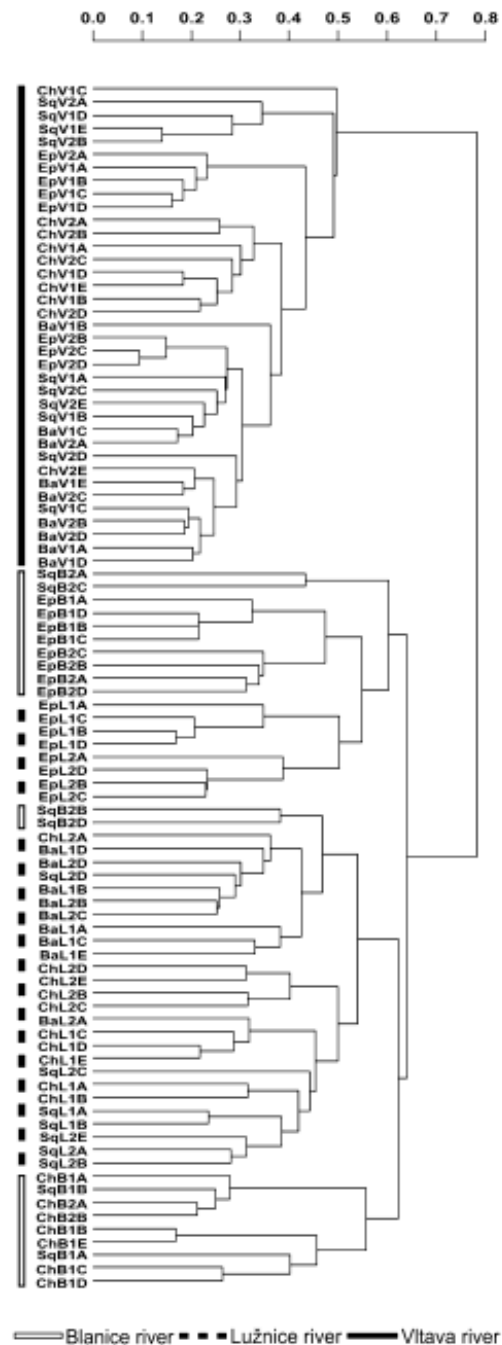
Tomáš Bešta · Milan Muška · Stephen Juggins · Jakub Těšitel

**Table 2** Hierarchical decomposition of variance in diatom communities sampled in the study

Hierarchical decomposition of variance			Mean square	<i>P</i>	Pairwise comparisons of Sample types		
Component	Explained variance (%)	df			Sample type	Explained variance (%)	Bonferroni <i>P</i> <sub>adj</sub>
River	47	2	23.5	0.066*	Ep-Ba	11.6	0.003
Locality	6.9	5	1.38	0.851**	Ep-Ch	17.3	0.003
Sample type	10	2	5	0.001	Ep-Sq	15.7	0.003
Residual	36.1	82	0.44	n.a.			

The proportions of variance are derived from a series of (partial) RDA analyses computed for each hierarchical level. The next higher hierarchical component was used as a covariate in analyses of variance accounted by Localities and Sample types. *P* value is based on a Monte-Carlo permutation test with 999 permutations. The species data were log-transformed prior analyses. Epilithic samples were excluded from hierarchical decomposition of variance, whereas they were used in pairwise comparisons of sample types. Pairwise comparisons of sample types are based on three partial RDAs (localities nested within rivers used as covariates) containing only samples relevant for the individual pairwise analyses. Due to the multiple testing, *P* value was adjusted using the Holm–Bonferroni method (Holm, 1979)

**Fig. 2** Cluster dendrogram based on Bray–Curtis dissimilarities constructed with using average linkage in package vegan. First two letters of the name are related to Sample type: *Ep* epilithon, *Ba* *Barbus barbus*, *Ch* *Chondrostoma nasus*, *Sq* *Squalius cephalus*. Second two letters are related to locality according Fig. 1



# OSNOVA PŘEDNÁŠKY

- Příprava dat pro numerické analýzy
  - typy sbíraných dat, čištění dat, odlehlé body, transformace, standardizace, EDA
- Design ekologických experimentů
  - manipulativní experimenty vs. pozorování
- Indexy druhové bohatosti
  - alfa, beta a gama diverzita, akumulární druhová křivka, *rarefaction*
- Ekologická nepodobnost
  - indexy nepodobnosti mezi vzorky
- Ordinance
  - lineární vs. unimodální, přímá vs. nepřímá, artefakty, ordinační diagramy, permutační testy, rozklad variance, parciální analýza, příkladové studie
- Klasifikace
  - hierarchická vs. nehierarchická, aglomerativní vs. divisivní

## SOFTWARE

- **R** – umí vše potřebné, zadarmo + kvalitní grafika, ovládání přes příkazový řádek
  - **Balíček Vegan (a další)**
- R – <https://cran.r-project.org/bin/windows/>



Jari Oksanen

- **CANOCO 5** – ordinační analýzy, diagramy, odpovědní křivky druhů, grafický interface



Cajo ter Braak



Petr Šmilauer



Jan „Šuspa“ Lepš



# LITERATURA

## V češtině

- Lepš J. & Šmilauer P. (2001) *Mnohorozměrná analýza ekologických dat*
  - v anglické verzi vyšlo v nakladatelství Cambridge v roce 2003 jako *Multivariate Analysis of Ecological Data using CANOCO 5* (v roce 2014 vyšlo druhé vydání pro CANOCO 5)
- Herben T. & Münzbergová Z. (2003) *Zpracování geobotanických dat v příkladech. Část 1. Data o druhovém složení*

## V angličtině

- Legendre P. & Legendre L. (2012) *Numerical Ecology* (Third English Edition). Elsevier.
- Lepš, Šmilauer (2014) *Multivariate analysis of Ecological Data using Canoco*. Cambridge University press. *Nejen pro uživatele Canoca*

## Webové zdroje

- David Zelený – *Analysis of community ecology data in R* <http://www.davidzeleny.net/anadat-r/doku.php/en:start>
- Palmer M. – *Ordination methods for ecologists*, website <http://ordination.okstate.edu/>

## R friendly

- Borcard D., Gillet F. & Legendre P. (2011) *Numerical Ecology with R*. Springer.

# ORGANIZAČNÍ INFORMACE KE KURZU

- Materiály k předmětu – postupně se budou objevovat v ISu
  - přednášky, příklady/data ke cvičení, studijní materiály
- Cvičení
  - probíhat bude v počítačové učebně a zaměřené bude na analýzu dat a jejich vizualizaci v programu **R**
- Domácí úkol
  - zadání bude sděleno v průběhu semestru
- Zkouška
  - Vypracování závěrečné práce - esej
  - Prezentace eseje na kolokviu

**TYPY SBÍRANÝCH DAT**  
**EXPLORAČNÍ ANALÝZA DAT**  
**ÚPRAVA DAT PRO ANALÝZU**

# MNOHOROZMĚRNÁ DATA

# DATA V EKOLOGII SPOLEČENSTEV

*Společenstvo je soubor druhů, které se vyskytují společně v prostoru a v čase.  
(Begon 2007)*

- **společenstvo** je studovaná (závislá) „proměnná“ (response variable)
- společenstvo je vícerozměrná „proměnná“ – zaznamenána v matici (data matrix)
- každý druh - jeho přítomnost nebo kvantita - představuje jeden rozměr společenstva
- zaznamenaný vzorek společenstva je reprezentativní pro určitou plochu a daný čas

(závislá, vícerozměrná proměnná nemusí nutně být tvořena druhy: mohou to být např. proměnné prostředí, vlastnosti druhů, atd.)

- prediktory (nezávislé proměnné) popisují nejčastěji **prostředí**, zkoumáme jejich efekt na strukturu společenstva

# USPOŘÁDÁNÍ DAT

## 1. druhová tabulka (matice)

- abundance, pokryvnosti, prezenze/absence druhů

## 2. proměnné prostředí

- hodnoty naměřených proměnných prostředí

## 3. (geografické proměnné)

- souřadnice lokalit

## 4. (popisné proměnné)

- další proměnné popisující vzorkovací schéma, např. studovaná oblast, determinátor organismů, vzorkující, datum/období odběru...

# TYPY PROMĚNNÝCH

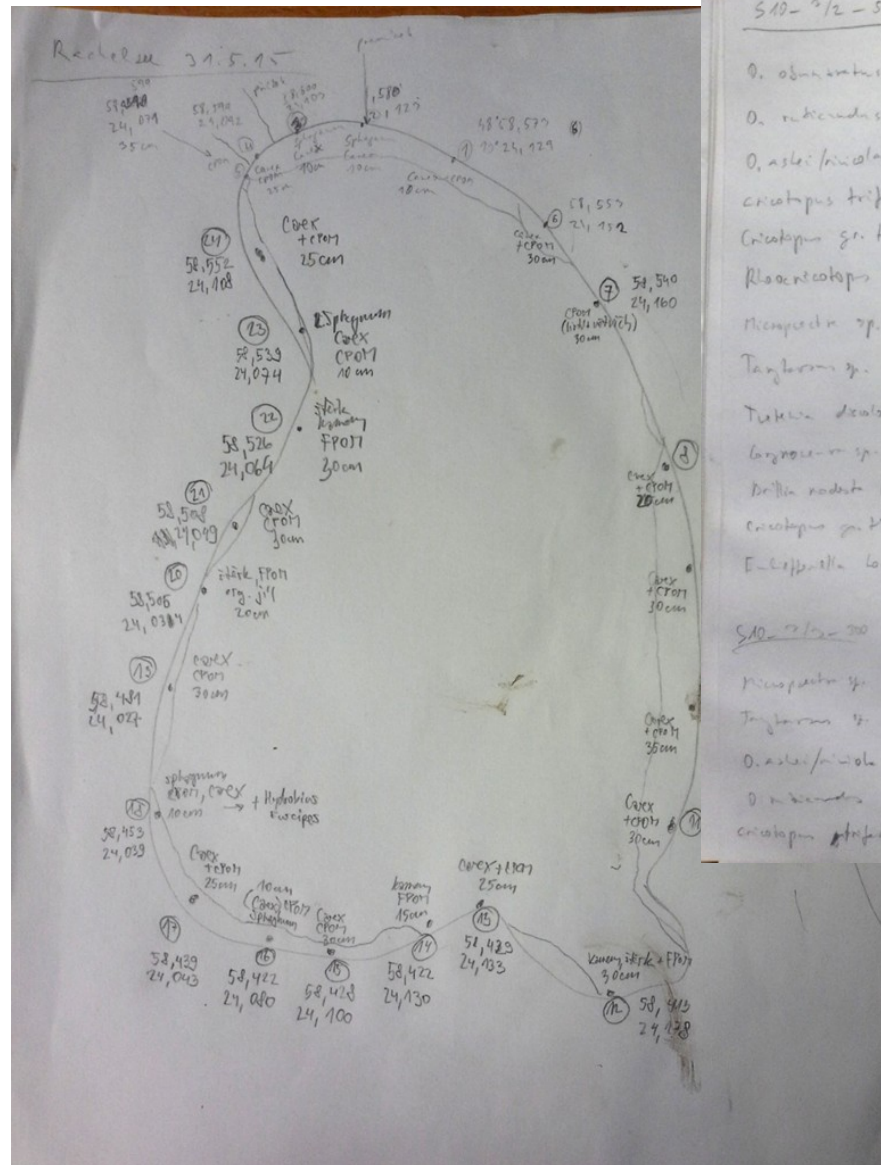
- **Kategoriální (kvalitativní, nominální, prezenčně-absenční)**
  - např. substrát, půdní typy, geografická oblast
  - binární proměnné (přítomnost-absence druhu)
  - kategorie jsou unikátní (každý jedinec/pozorování spadá právě do jedné z nich)
  - kategorie nelze smysluplně seřadit
- **Ordinální (semikvantitativní)**
  - např. Ellenbergovy indikační hodnoty pro druhy, Braun-Blanquetova stupnice pro odhad pokryvnosti druhů
  - jednotlivé stupně (kategorie) lze seřadit, rozdíly mezi sousedními stupni jsou různě velké
- **Kvantitativní**
  - **diskrétní** (počty jedinců, měření s malou přesností), rozdíly mezi sousedními stupni jsou stejně velké
  - **kontinuální** (přesná měření, např. množství biomasy,)

## KVANTITATIVNÍ VS. SEMIKVANTITATIVNÍ

- kvantitativní sice přesnější
- ale semikvantitativní rychlejší a levnější
- trade-off mezi počtem vzorků a přesností
  - Počet vzorků obvykle zvyšuje vypovídací hodnotu analýzy podstatně více než přesnost dat (zákon velkých čísel)
- semikvantitativní často postačující
  - Neboť statistika je přece soubor přesných metod, popisujících, jak zacházet s nepřesnými daty



# PRIMÁRNÍ DATA - SBĚR



S10- 2/2 - 500 Synanthel. sem S.d. 9 ✓

O. oblongatus IIII IIII IIII ✓ 1 ✓  
 O. rubicundus IIII IIII IIII IIII IIII ✓ 1 ✓  
 O. asclepioides IIII IIII IIII ✓  
 Cricotopus trifasciatus ✓  
 Cricotopus gr. hirsutus ✓  
 Rhacocricotopus fuscipes ✓  
 Micropsectra sp. II ✓  
 Tanytarsus sp. I ✓  
 Tursina deulaplesi formosa S.d. IIII ✓  
 Gynostomum sp. II ✓  
 Drilina nodata I ✓  
 Cricotopus gr. hirsutus II ✓  
 Ectopsectra lobifera S.d. ✓

S10- 2/2 - 500 Synthel. an. 24 III

Micropsectra sp. 9 ✓  
 Tanytarsus sp. I ✓  
 O. asclepioides IIII IIII ✓  
 O. rubicundus IIII IIII IIII IIII IIII ✓  
 Cricotopus trifasciatus IIII ✓

Podatý:

S10- 2/2 - 500 1L - Cricotopus cruetus  
 1P - P. ruficornis, C. hirsutus ✓  
 S10- 3/2 - 500 4L - P. ruficornis ✓ hirsutus v. 1 postoj. per  
 S10- 2/2 - 500 5P - C. gr. hirsutus ✓  
 S10- 2/2 - 500 1P - C. asclepioides ✓  
 S10- 2/2 - 500 (5P) - C. cruetus ✓  
 S10- 4/2 - 500 (2P) - Cricotopus trifasciatus ✓ (S.d. 2/2)  
 hirsutus s. C. trifasciatus S10- 2/2 - 500 (224)  
 drilina ✓  
 20.11.95 v. trifasciatus ✓  
 h. hirsutus ✓

Ideálně ihned po návratu naskenujeme a pečlivě uložíme papírovou i digitální verzi

# PRIMÁRNÍ DATA – PŘEPIS

- tabulkový editor (spreadsheet)
  - např. Microsoft Excel
- přepisujeme co nejdřív
  - zachycení chybějících dat
  - v poznámkách se snáz orientujeme, po čase přestanou být srozumitelné
  - dvě kopie je obtížnější ztratit nebo zničit
  - přítomnost dat v počítači podpoří jejich rychlé zpracování

- metadata – data o datech

- jméno studie
- kdo a kde data sbíral, za jakým účelem
- popis experimentálních jednotek, objektů
- metodika sběru dat a měření (velmi oceníme při psaní práce)
- popis proměnných, jednotky
- popisy zkratk

EXIF Information			
File name:	DSC_0260.JPG	File size:	922866 bytes
File date:	2006:04:22 22:06:16	Camera make:	NIKON CORPORATION
Camera model:	NIKON D70s	Date/Time:	2006:04:17 18:06:08
Resolution:	3000 x 2632	Flash used:	No
Focal length:	18.0mm (35mm equivalent: 27mm)	Exposure time:	0.0008 s (1/1250)
Aperture:	f/8.0	Whitebalance:	Manual
Metering Mode:	matrix	Exposure:	Manual
Exposure Mode:	ManualAuto bracketing		

# PRIMÁRNÍ DATA – SPREADSHEET

- v řádcích studované objekty
  - Vzorky (*samples*)
- v sloupcích proměnné, kterými objekty charakterizujeme
  - Druhy (*species*), proměnné prostředí (*predictors, descriptors*)

	A	B	C	D
1	FIELD_NO	Deroceras altaicum	Deroceras laeve	Euconulus fulvus
2	IA003			5
3	IA004			
4	IA005		1	5
5	IA006			
6	IA007		2	
7	IA008		3	
8	IA010			5
9	IA011			2
10	IA013			
11	IA015			3
12	IA017			
13	IA018			
14	IA022			

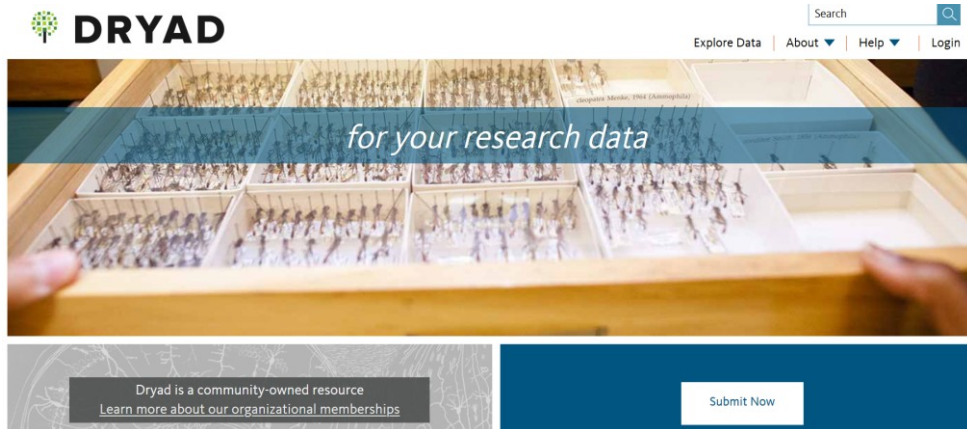
	A	B	C	D	E	F	G	H
1	FIELD_NO	RICHNESS	VEGEALTA	deg_lat	deg_lon	PREC_1	PREC_7	TMEAN_1
2	170-06	95	Forest	51.50142	86.01867	13	74	-14.3
3	171-06	107	Forest	51.53725	85.97119	13	74	-14.3
4	172-06	114	Forest	51.52756	85.96128	13	74	-14.3
5	JD201	43	Open	50.47675	87.62292	9	69	-19.6
6	JD202	18	Forest	50.47717	87.60958	9	69	-19.8
7	JD203	38	Forest	50.47833	87.60978	9	69	-19.8
8	JD204	39	Open	50.48056	87.61336	9	69	-19.8
9	JD205	53	Open	50.47622	87.61264	9	69	-19.8
10	JD206	61	Open	50.70264	88.01653	7	67	-17.8
11	JD207	38	Open	50.49731	87.65717	9	69	-19.8
12	JD208	34	Open	50.49619	87.65397	9	69	-19.8
13	JD209	44	Forest	50.43617	87.58992	9	68	-19.3

vysvětlivky k proměnným

	A	B	
1	FIELD_NO	ID	Hlavní identifikace
2	RICHNESS-PLANTS	Závislá proměnná: počet druhů c	
3	VEGEALTA2	Klasifikátor datasetu na podsouk	
4	deg_lat	Koordináty	Latitude, formát
5	deg_lon	Koordináty	Longitude, formát
6	PREC_1	Macroclimate	Srážky leden
7	PREC_7	Macroclimate	Srážky červenec
8	TMEAN_1	Macroclimate	Teplota leden
9	TMEAN_7	Macroclimate	Teplota červenec
10	BIO_1	Macroclimate	BIO1 = Annual Mean Temperature
11	BIO_2	Macroclimate	BIO2 = Mean Diurnal Range (Mean of monthly (max temp -

# ZÁLOHA DAT

- Uchování a zpřístupnění primárních dat
  - problematika dlouhodobé archivace a nosičů dat (nejlepší je stále papír bez volných kyselin + laserová tiskárna)
  - zpřístupnění primárních dat (některé časopisy, např. *Ecological Monographs*, *Journal of Ecology* aj., to mají jako podmínku zveřejnění článku)
  - uložení dat ve veřejně dostupných elektronických repozitářiích (např. Dryad Digital Repository, [www.datadryad.org](http://www.datadryad.org)) nebo databázích (např. Česká Národní Fytocenologická Databáze)



# KONTROLA DAT

## ○ chyby (*errors*)

- někdy se chovají jako odlehlé body, je třeba zkontrolovat původní záznam a případně data opravit
- Velmi omezená možnost kontroly dat zadávaných přímo digitálně (třeba do telefonu)

## ○ chybějící data (*missing data, NA*)

- Mnohorozměrné metody typicky netolerují NAs
- možnosti jejich nahrazení (interpolace, model)
- vyloučení proměnné nebo vzorku který má hodně chybějících hodnot

## ○ odlehlé body (*outliers*)

## ○ další úpravy:

- sjednocení taxonomické nomenklatury
- taxonomická adjustace (např. kombinace dvou nerozlišitelných druhů)

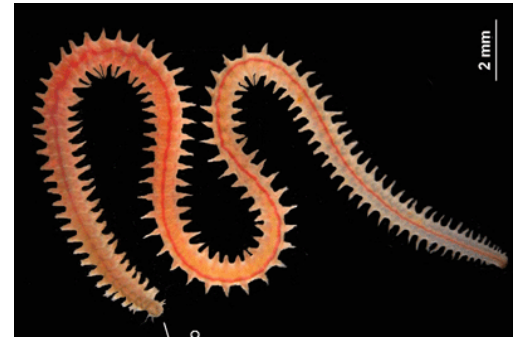
	W	X	Y	Z	AA	AB
1 PLOT_ID	MO-03	MO-04	MO-05	MO-06	MO-07	MO-09
2 DATE	#####	#####	#####	#####	#####	#####
3 TIME OF START	17:35	16:00	15:07	13:40	12:40	17:21
4 AUTHORS	David (ori)	David	David	David	David	Markéta+
5 Fisheye_1	7098	7064	7117	7114	7392	8131
6 Fisheye_2	7097		7118	7113	7393	8130
7 Fisheye_3	7096	7066	7119	7112	7394	8127
8 Fisheye_4	7065	7067	7120	7111	7395	8128
9 Fisheye_5	7094	7068	7121	7110	7396	8129
10 Fisheye_6	7093	7069	7122	7109	7397	8126
11 Fisheye_7	7092	7070	7123	77108	7398	8125
12 Fisheye_8	7091	7071	7124	7107	7399	8124
13 Soil_1	13	3	18	10	10	0
14 Soil_2	3	24	20	8	10	1
15 Soil_3	11	5 30+		0	9	13
16 Soil_4	10	1	29	11	14	2
17 Soil_5	7	5	11	13	3	10
18 Soil_6	4	9		4	6	1
19 Soil_7	7	10	14	11	3	7
20 Soil_8	1	4	19	0	11	14
21 Soil_9	10	0	18	11	6	2
22 Soil_10	0	1	7	2	6	4
23 Soil_11	23	22	8	4	9	3
24 Soil_12	10	0	2	14	15	10
25 Soil_13	12	9	11	9	13	17
26 Soil_14	3	0	9	9	11	9
27 Soil_15	10	10	10	3	12	2
28 Soil_16	5	3	3		11	6
29 Soil_17	1	19	20	0	5	3
30 Soil_18	9	15	9	24	6	2
31 Soil_19	17	12	17	10	16	11
32 Soil_20	8	16	7	12	14	14
33 Soil_21	7	12	2	8	9	13
34 Horiz_0	4	6	9	2	6	19
35 Horiz_45	16	4	6	14	4	19
36 Horiz_90	0	0	3	0	2	5
37 Horiz_135	0	4	0	7	0	11
38 Horiz_180	3	5	0	0	0	4
39 Horiz_225	0	4	0	8	0	5

# EXPLORATORY DATA ANALYSIS (EDA)

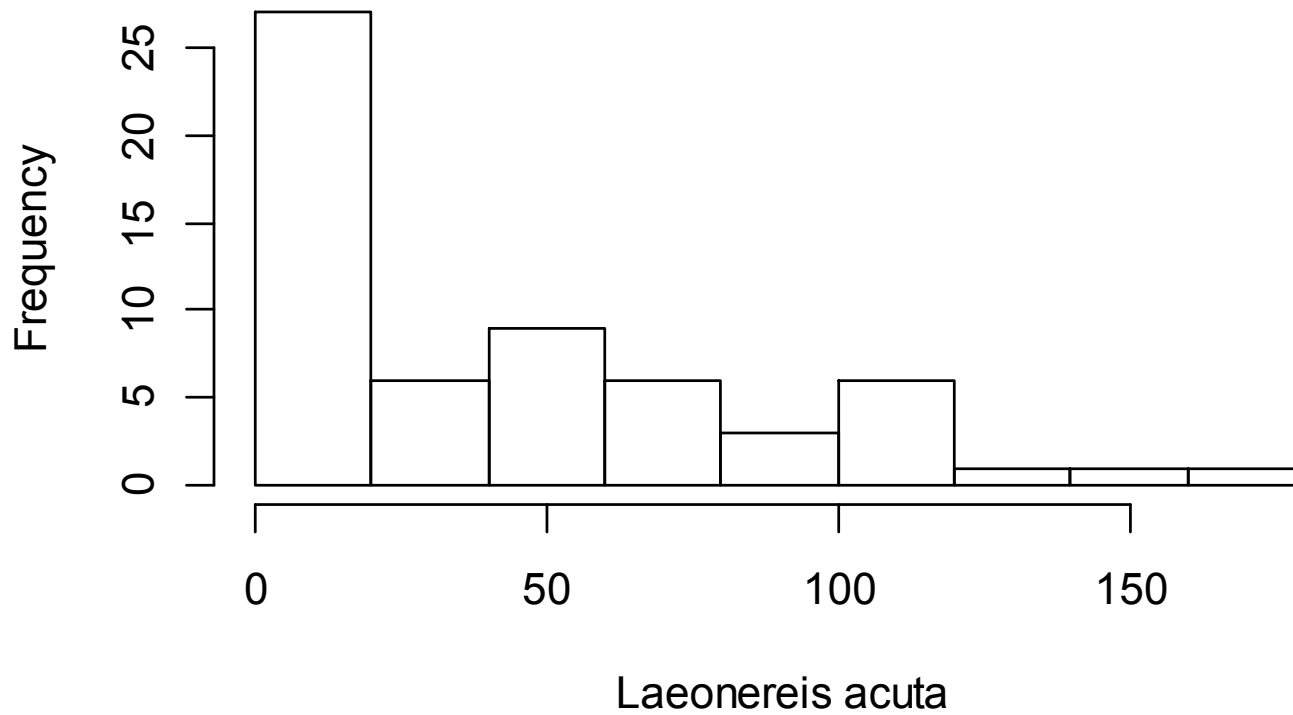
Proč EDA:

- odhalení odlehlých bodů (*outliers*)
- prozkoumání rozložení hodnot
  - střední hodnota
  - rozsah
  - tvar rozložení (normalita, bimodalita, sešikmenost)
- odhalení vztahů mezi proměnnými
- posouzení nutnosti transformace
- měla by zabrat ~20% celkového času studie

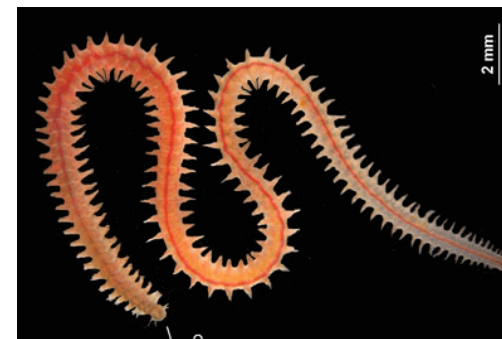
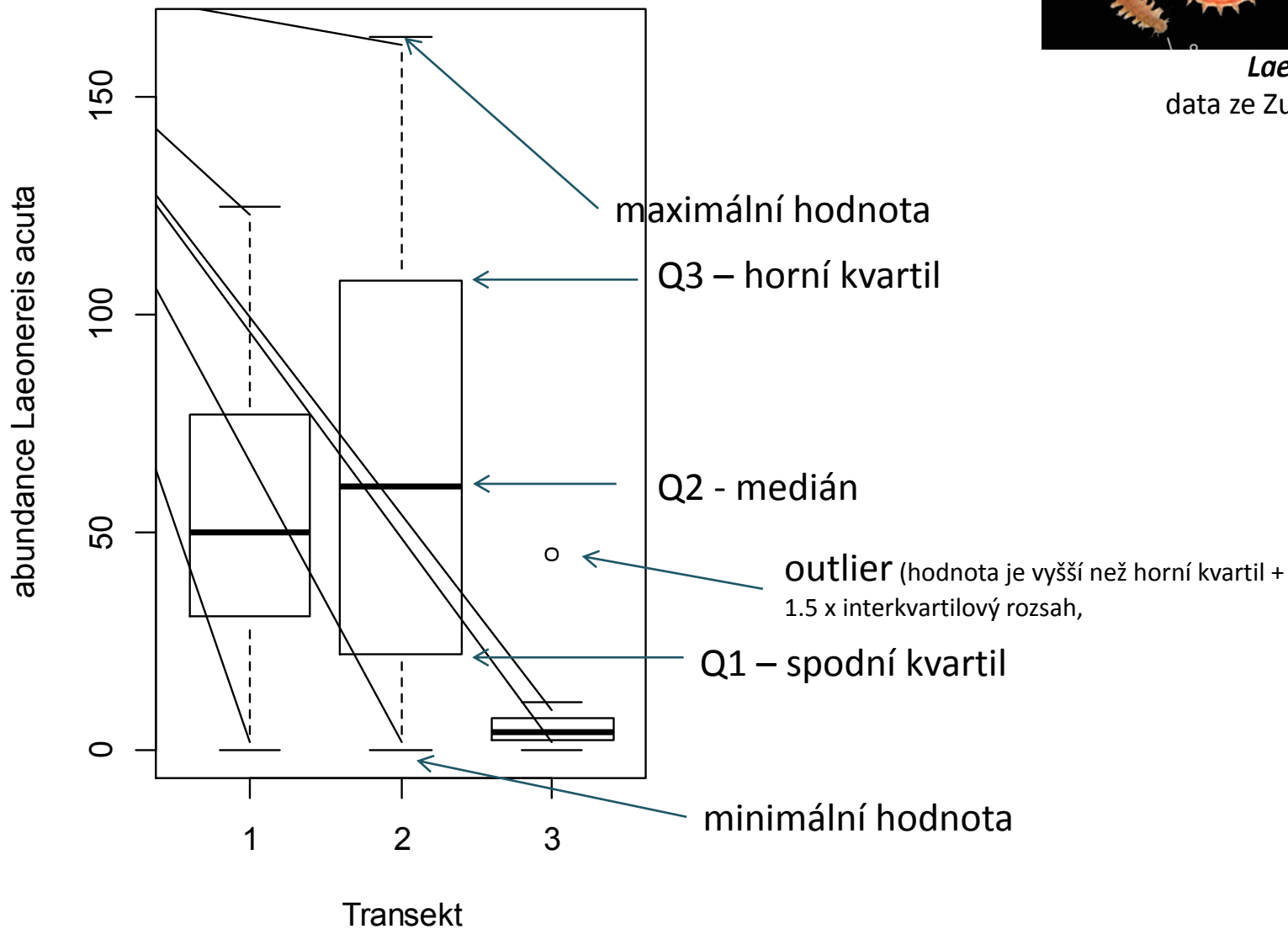
# HISTOGRAM



*Laeonereis acuta*



# KONDITIONÁLNÍ BOXPLOT

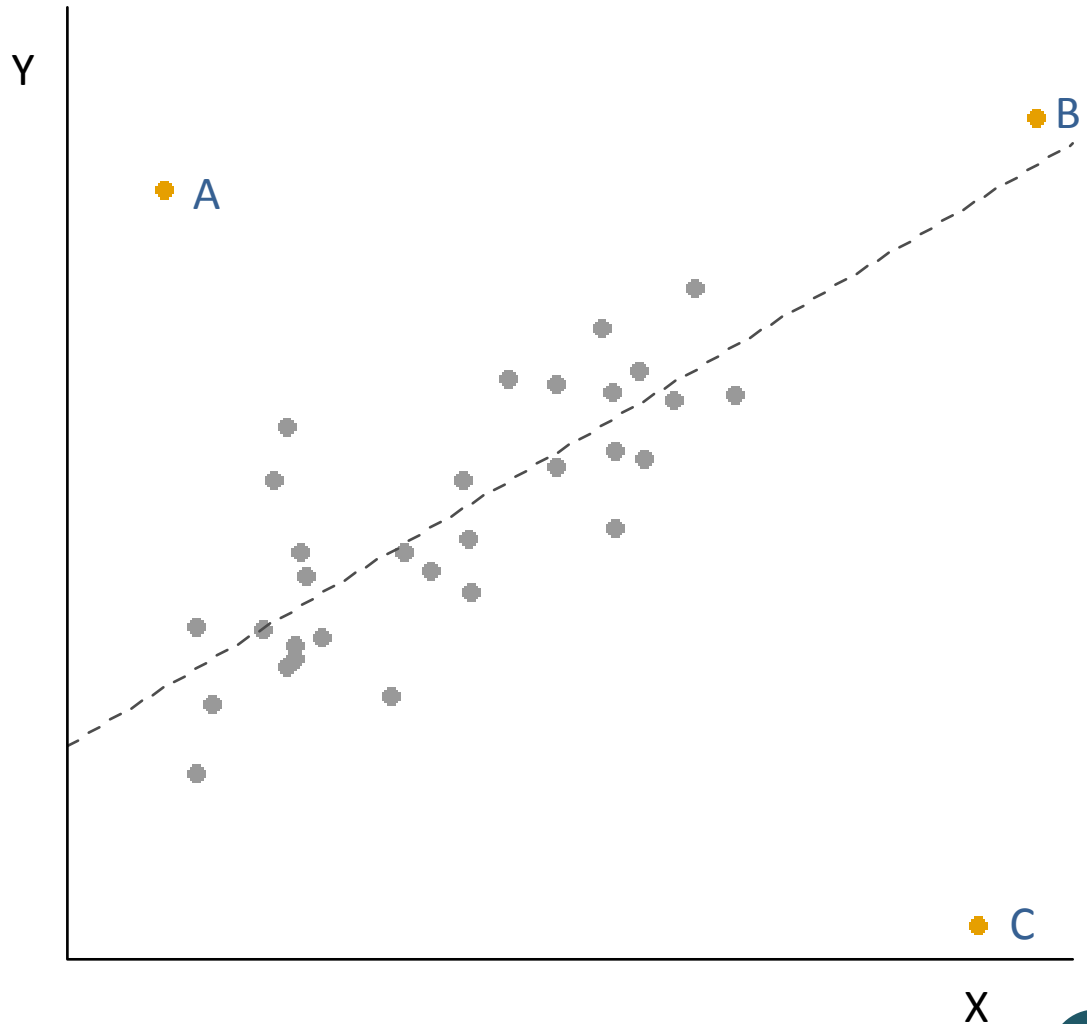


*Laeonereis acuta*  
data ze Zuur et al. 2007



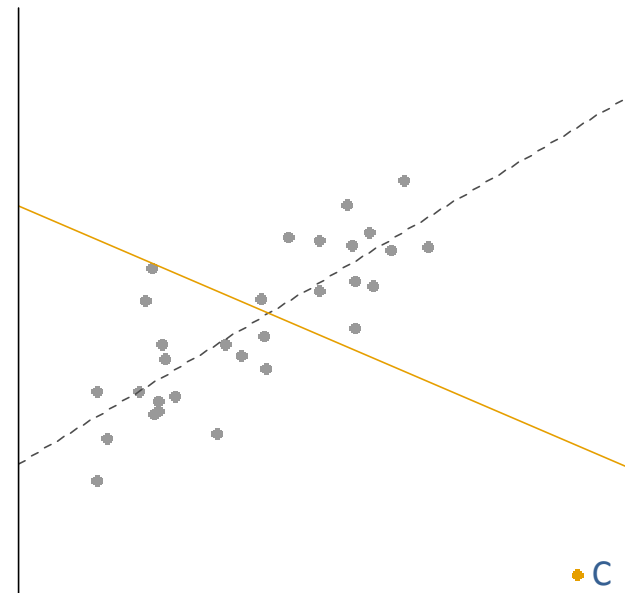
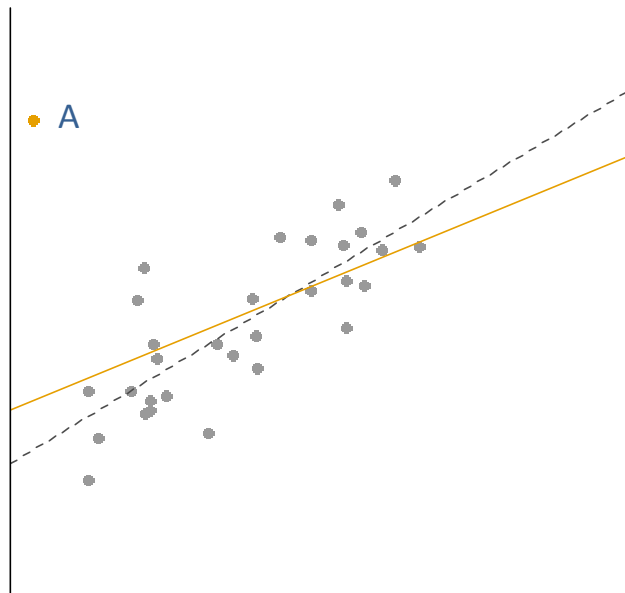
# BODOVÝ GRAF (SCATTER PLOT)

- A. outlier z hlediska vztahu Y a X
- B. outlier z hlediska rozložení hodnot X i Y
- C. outlier z hlediska rozložení hodnot i vztahu Y a X

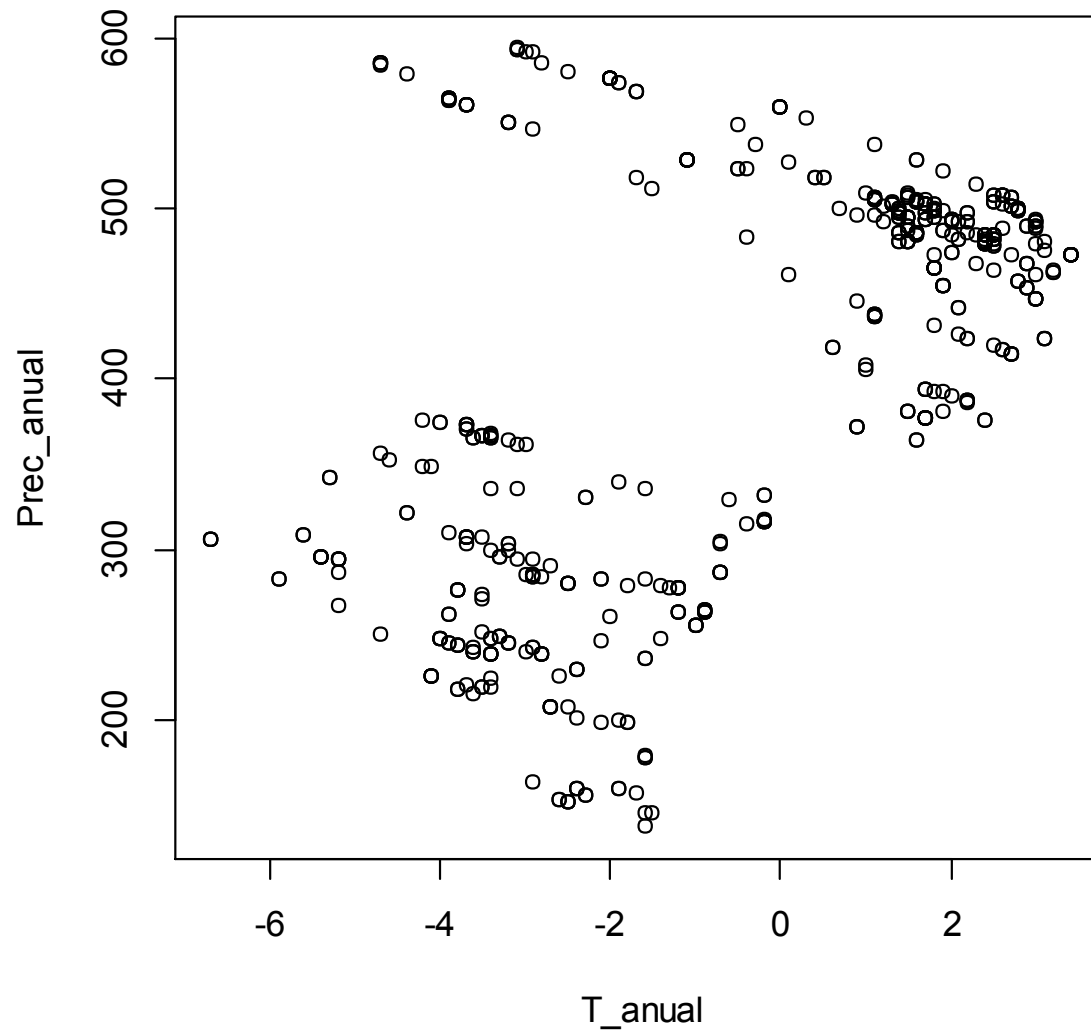


# BODOVÝ GRAF (SCATTER PLOT)

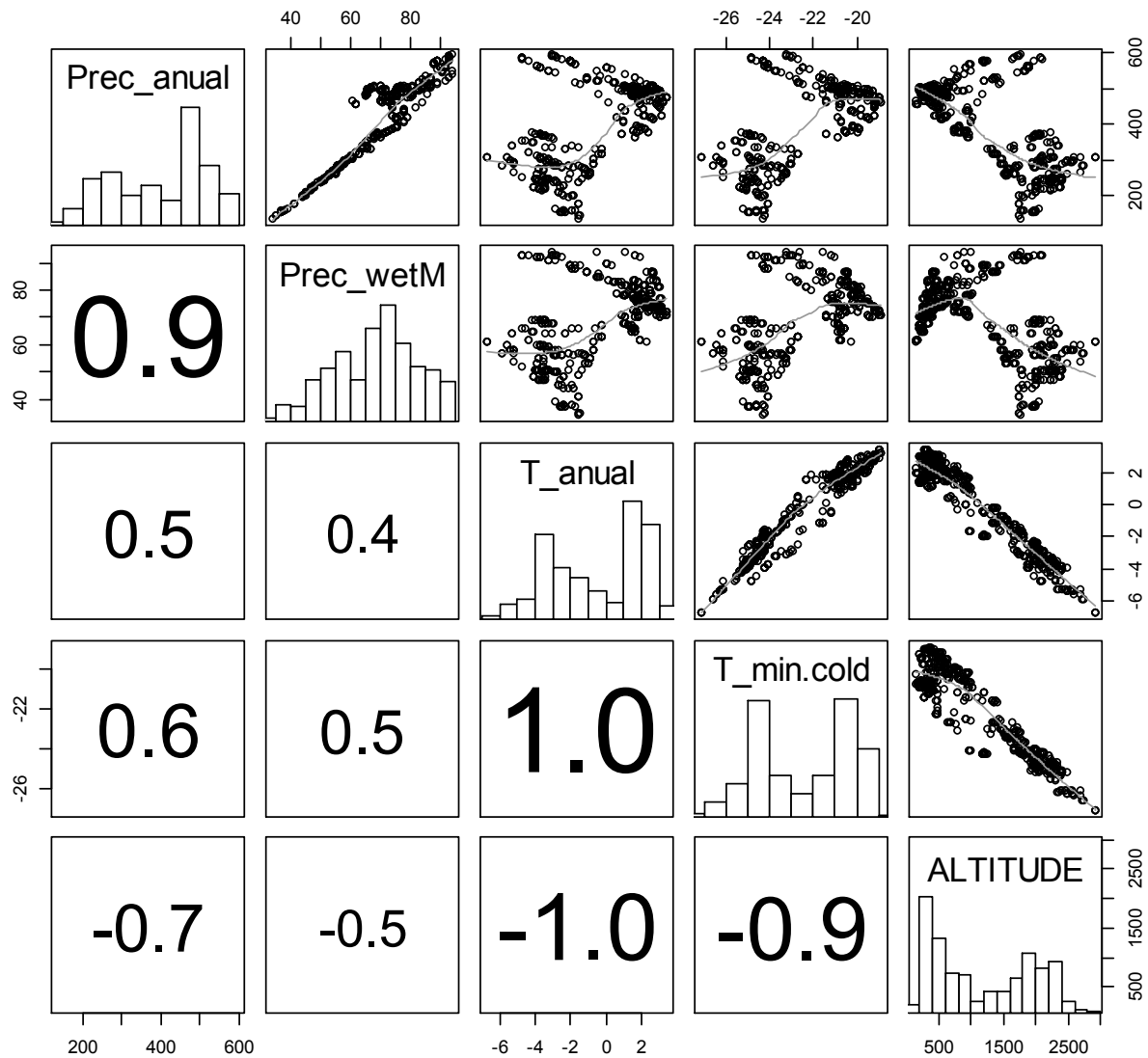
- dramatický vliv outlierů na výsledky analýz



# BODOVÝ GRAF (SCATTER PLOT)



# PÁROVÝ GRAF (PAIR PLOT)



## CO S OUTLIERY?

- automatické odstranění z datasetu – **špatně!**
- odstranění pouze když:
  - data chybně zaznamenána
  - nespádají do zamýšleného studovaného prostoru (*sample space*), např. v důsledku havárie čističky nad lokalitou, nebo louka rozježděna čtyřkolkami
- některé body se jeví jako outliery jen proto, že je nutíme do normálního rozložení
  - extrémy v datech z log-normálního nebo exponenciálního rozložení po transformaci krásně zapadnou mezi ostatní