

# ZPRACOVÁNÍ DAT V EKOLOGII SPOLEČENSTEV

JAKUB TĚŠITEL S POUŽITÍM MATERIÁLŮ VÍTA SYROVÁTKY A DAVIDA ZELENÉHO

# Dipteran assemblages of spring fens closely follow the gradient of groundwater mineral richness

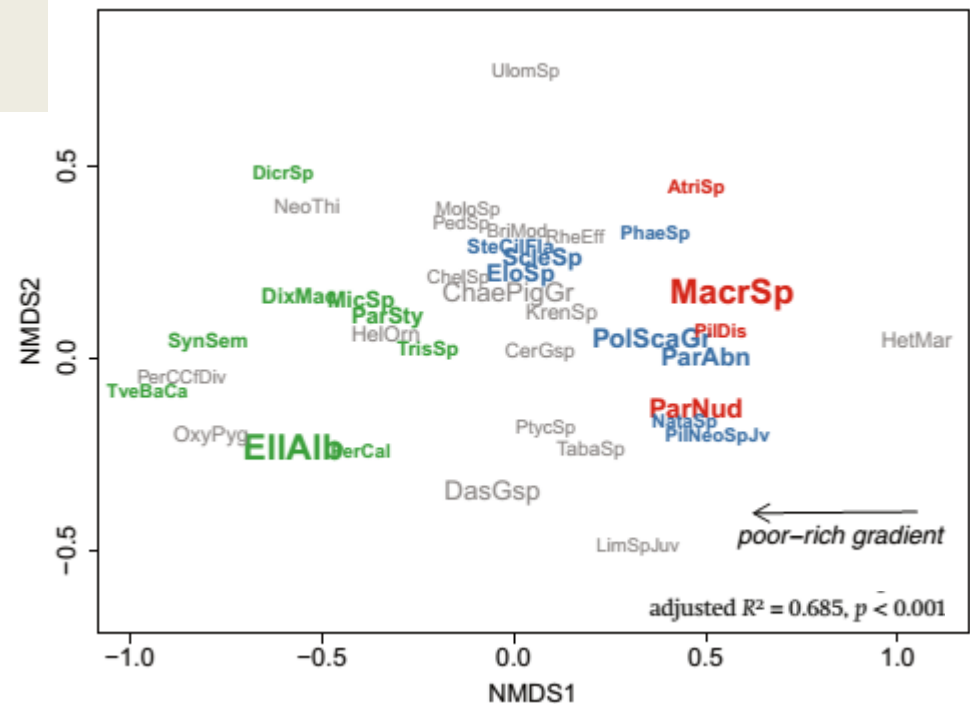
Markéta Omelková, Vít Syrovátka, Vendula Křoupalová, Vanda Rádková, Jindřiška Bojková, Michal Horsák, Marie Zhai, and Jan Helešic

Can. J. Fish. Aquat. Sci. 70: 689–700 (2013) [dx.doi.org/10.1139/cjfas-2013-0026](https://doi.org/10.1139/cjfas-2013-0026)

- Společenstvo larev dvoukřídlých na prameništích slatiništích.
- ? Mění se podél gradientu minerální bohatosti (obdobně jako společenstva jiných organismů, např. cévnatých rostlin)?
- ? Jaké další faktory určují skladbu společenstva?

**Table 2.** The results of PERMANOVA analysis indicating the importance of environmental variables for the structuring of dipteran assemblages within the flowing water and standing water mesohabitats, respectively.

	Variable	F	R <sup>2</sup>	p
Flowing water	Poor–rich gradient	3.818	0.184	<0.001
	D <sub>50</sub>	2.030	0.098	0.009
	Temperature	1.899	0.092	0.022



# 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

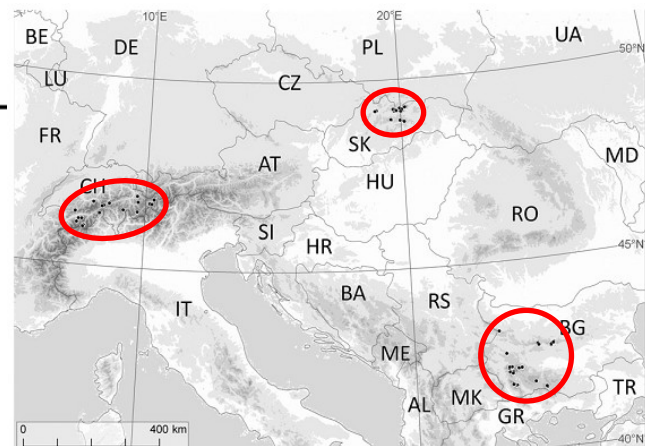


Fig. 1 The location of vegetation plots in the study regions. The country codes are indicated

- Vegetace podhorských pramenů ve:
  - Švýcarských Alpách,
  - Bulharských horách
  - Západních Karpatech
- ? Liší se  $\beta$ -diverzita mezi horskými oblastmi?

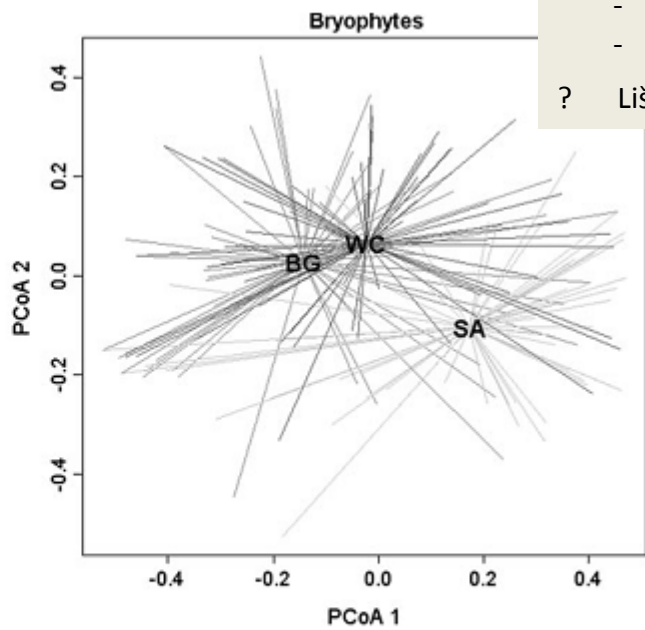
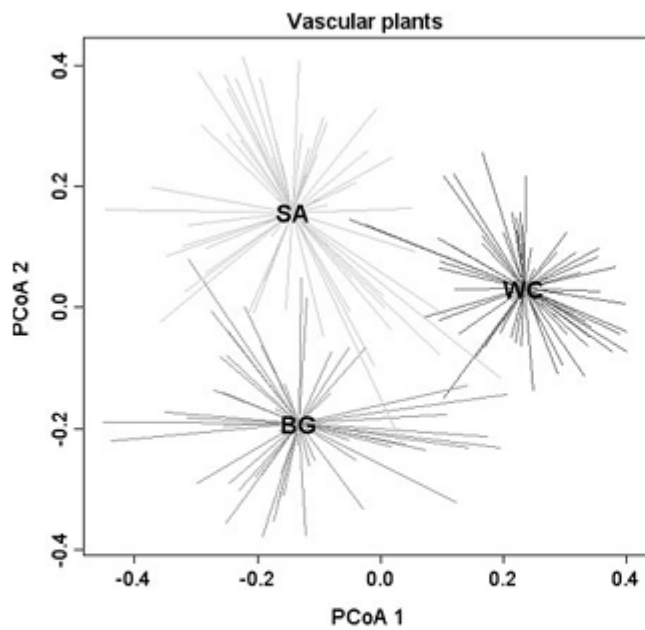


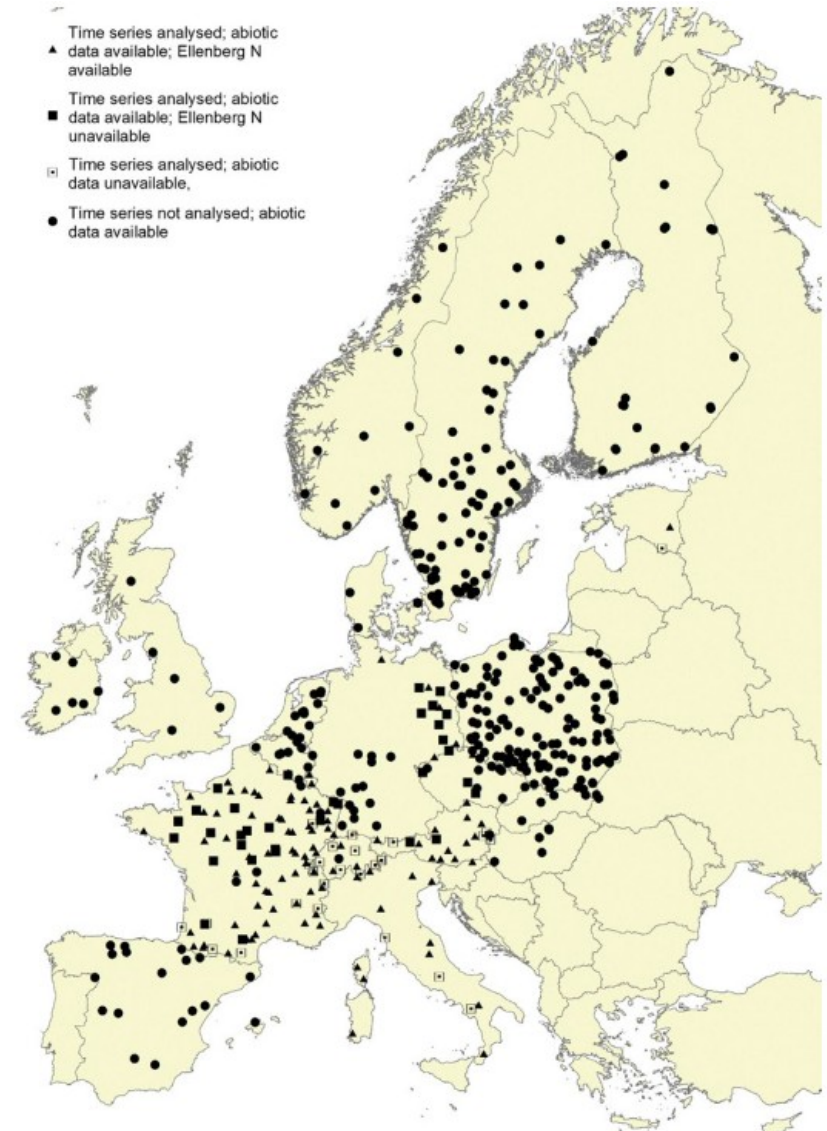
Fig. 4 Comparison of overall  $\beta$ -diversity of spring vegetation in the three study regions for vascular plants and bryophytes.  $\beta$ -Diversity (dispersion of groups) of vascular plants was highest in Swiss Alps and lowest in Western Carpathians while no

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

## The contribution of nitrogen deposition to the eutrophication signal in understorey plant communities of European forests

Han F. van Dobben  | Wim de Vries

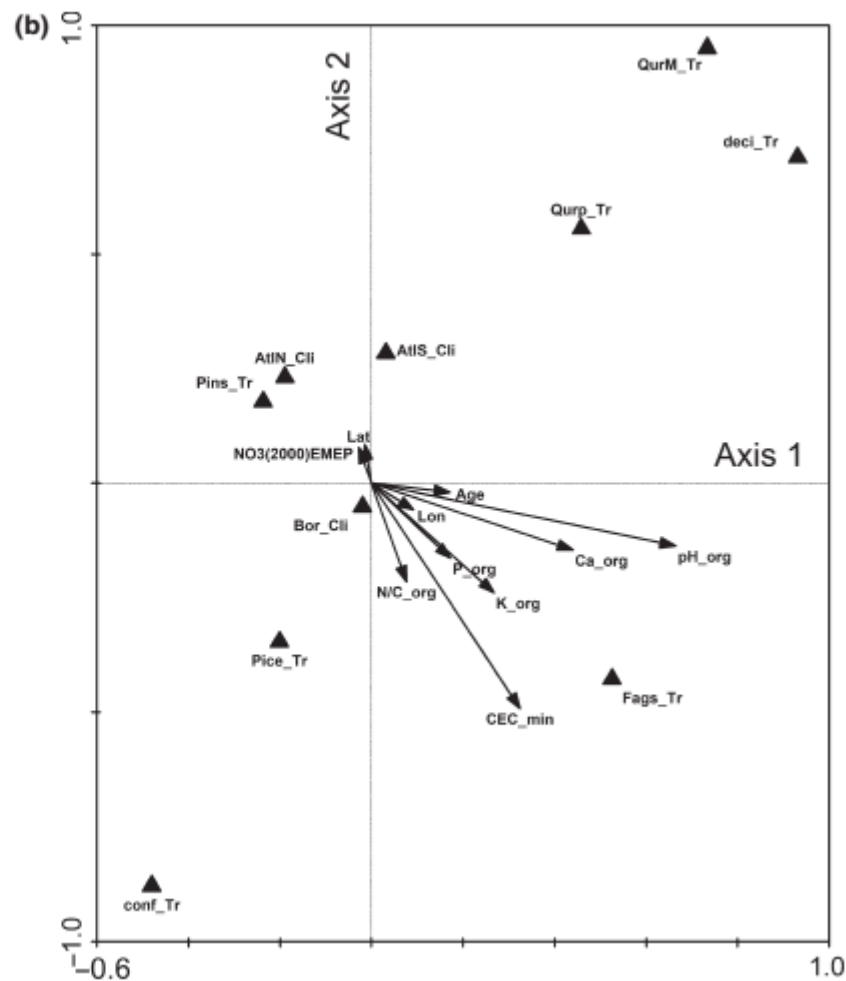
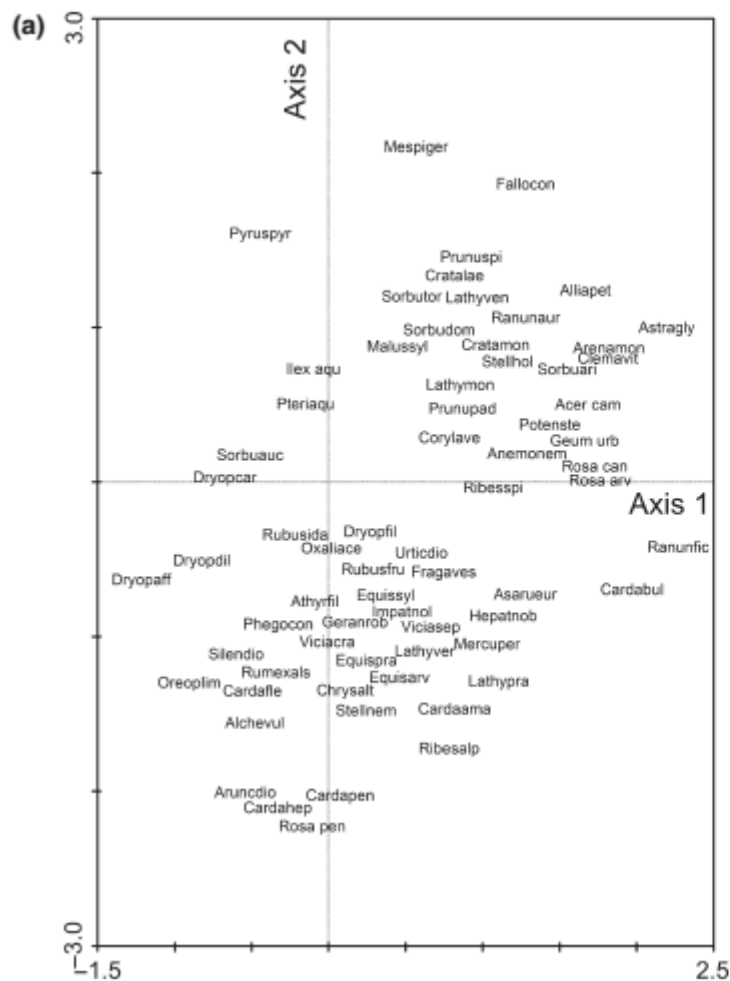
- Analýza vztahu mezi vegetací a prostředím
- Proměnné prostředí:
  - tree species
  - climate zones
  - altitude
  - geographical coordinates
  - stand age
  - soil chemical variables
  - deposition estimates of NO<sub>3</sub>, NH<sub>4</sub>, and N-total
- ? Jaké faktory zejména určují druhovou skladbu rostlinných společenstev?
- ? Jaké další faktory určují skladbu společenstva?



# The contribution of nitrogen deposition to the eutrophication signal in understory plant communities of European forests

Han F. van Dobben | Wim de Vries

Compartment	% Explained variance (data)	% Explained variance (fitted values)
Tree layer	5.9	47.3
Soil organic layer	3.4	27.4
Soil mineral layer	0.6	4.8
Climate	2.0	16.4
Deposition	0.5	4.1
Sum	12.4	100.0



# Drought mediates the importance of stochastic community assembly

Jonathan M. Chase\*

Department of Biology, Washington University, Box 1229, St. Louis, MO 63130

Edited by James H. Brown, University of New Mexico, Albuquerque, NM, and approved September 12, 2007 (received for review May 9, 2007)

Historically, the biodiversity and composition of species in a locality was thought to be influenced primarily by deterministic factors. In such cases, species' niches create differential responses to environmental conditions and interspecific interactions, which

stave off invasion by subsequent colonizing species through priority effects leading to multiple stable states (11). In contrast, "niche-selection" (3) predominates if species' abilities to establish in a locality are more strongly determined by their traits.

- 4-letý experiment
  - oživení malých nádrží
  - náhodná introdukce druhů
  - po 2 letech vypuštění poloviny nádrží (drsné prostředí, ekologický filtr)
  - po dalších 2 letech vzorkování
- ? Jsou společenstva drsného prostředí stejně rozmanitá?

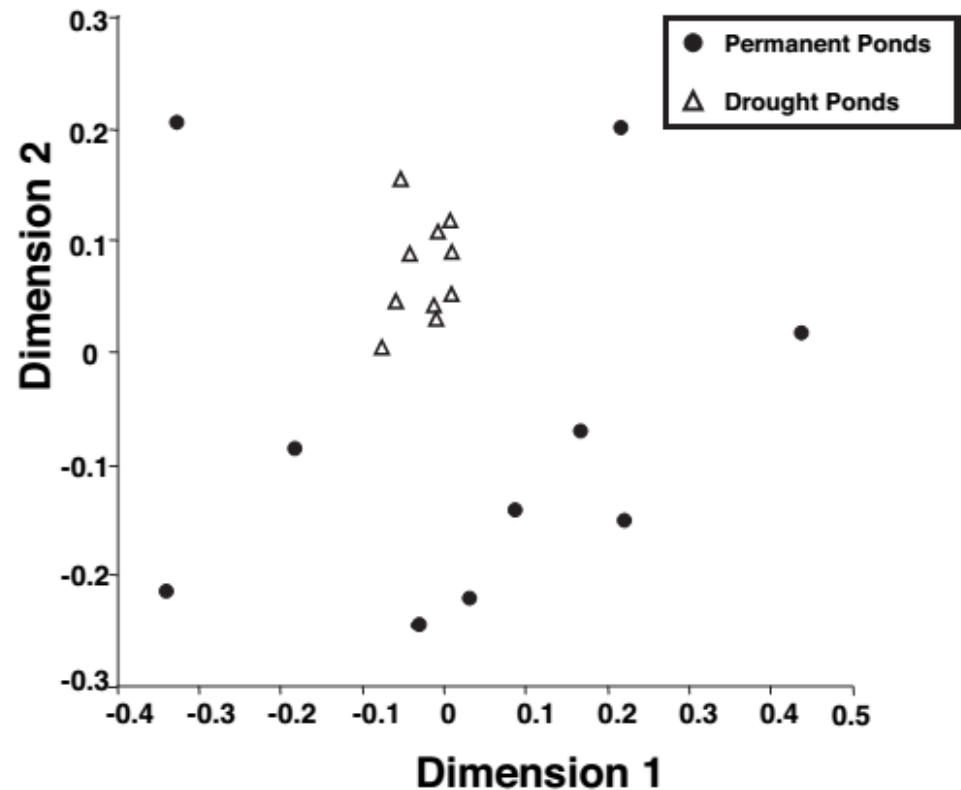


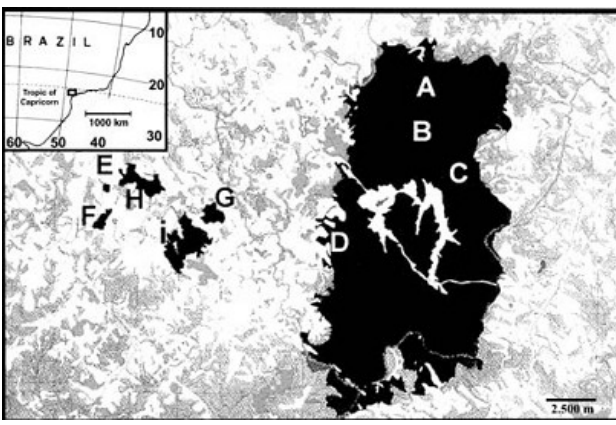
Fig. 2. Nonmetric multidimensional scaling (NMDS) allows the visualization of the multidimensional composition of communities in two-dimensional space. The Euclidean distance between any two points represents the difference in Jaccard's similarity between those two communities. Drought ponds are indicated by open triangles, and permanent ponds are indicated by filled circles.



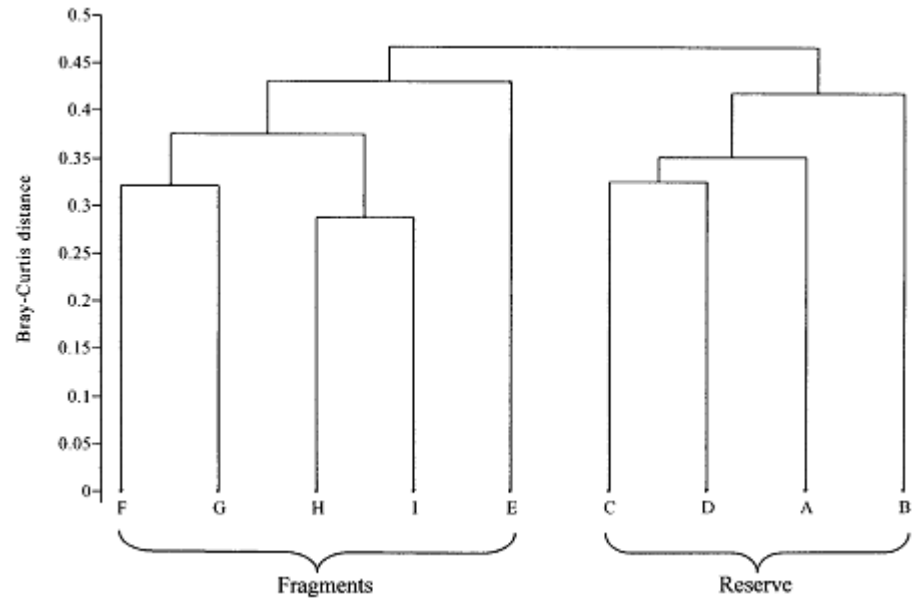
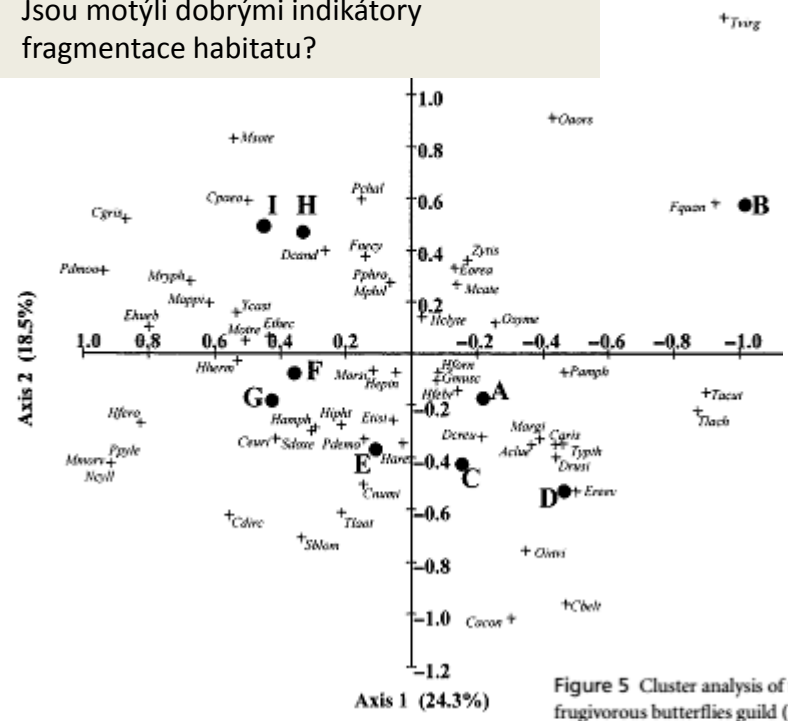
**RESEARCH PAPER**

# Species richness, composition and abundance of fruit-feeding butterflies in the Brazilian Atlantic Forest: comparison between a fragmented and a continuous landscape

Marcio Uehara-Prado<sup>1,2\*</sup>, Keith Spalding Brown, Jr<sup>1</sup> and André Victor Lucci Freitas<sup>1</sup>



- společenstvo motýlů v Brazílském pralese
  - velká plocha vs. fragmentovaný habitat
- ? Jsou motýli dobrými indikátory fragmentace habitatu?



**Figure 5** Cluster analysis of the sampling units of Morro Grande Reserve (A–D) and in the fragmented landscape (E–I) (UPGMA) of the frugivorous butterflies guild (above). Correspondence analysis of the same data set (below). The original data set was Hellinger-transformed. Code names are the first letter of the genus and the four first letters of the species (except *Pdemoo* = *Prepona demophoon*), following the species list in Appendix 1 in Uehara-Prado *et al.* (2005).

# OSNOVA PŘEDNÁŠKY

- Příprava dat pro numerické analýzy
  - typy sbíraných dat, čištění dat, odlehlé body, transformace, standardizace, EDA
- Indexy druhové bohatosti
  - alfa, beta a gama diverzita, akumulární druhová křivka, *rarefaction*
- Ekologická podobnost
  - indexy podobnosti a vzdálenosti 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, řízená vs. neřízená
- Design ekologických experimentů
  - manipulativní experimenty vs. přírodní experimenty (pozorování)
- Případové studie na použití jednotlivých metod



## 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 (pokyny viz webová stránka předmětu, sekce *Závěrečná práce*)
  - zhruba půlhodinová diskuze nad závěrečnou prací, doplněná o rozšiřující otázky týkající se probírané látky

**TYPY SBÍRANÝCH DAT**

**EDA**

**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

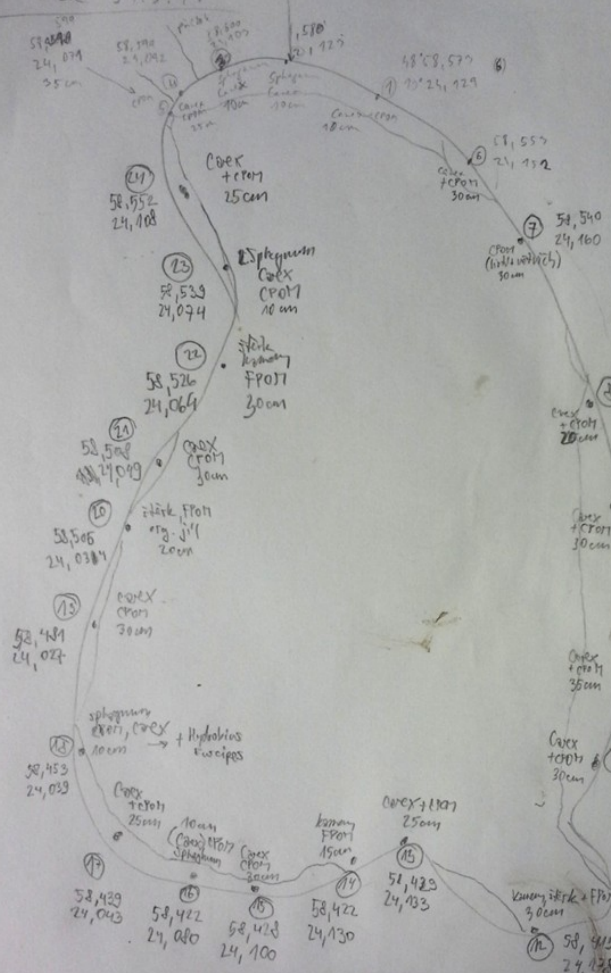
- **Kategoriální (kvalitativní, nominální, prezenčně-absenční)**
  - např. substrát, půdní typy, geografická oblast
  - binární proměnné (přítomnost-absence druhu)
  - kategorie jsou unikátní (každý jedinec/pozorování spadá právě do jedné z nich)
  - kategorie nelze smysluplně seřadit
- **Ordinální (semikvantitativní)**
  - 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)

## KVANTITATIVNÍ VS. SEMIKVANTITATIVNÍ

- kvantitativní sice přesnější
- ale semikvantitativní rychlejší a levnější
- trade-off mezi počtem vzorků a přesností
- semikvantitativní často postačující

# PRIMÁRNÍ DATA - SBĚR

Racholice 31.5.95



S10- 2/2 - 500 Spornábel. sem s.d. 9 v

- O. obscurus IIII IIII IIII I v
- O. rubicundus IIII IIII IIII IIII IIII v
- O. aster/nivola IIII IIII IIII v
- Cratopus triplicatus I v
- Cratopus gr. hirsutus I v
- Rhaenocotaps furcipes I v
- Micropsecta sp. II v
- Tanytarsus sp. I v
- Turbina dealbata/umbell. s.d. IIII v
- Garynesia sp. II v
- Dethia nodata I v
- Cratopus gr. hirsutus II v
- E-leptopella lobifera s.d. I v

S10- 2/2 - 500 Spornábel. sem. 24 III

- Micropsecta sp. 9 v
- Tanytarsus 4. I v
- O. aster/nivola IIII IIII -
- O. rubicundus IIII IIII IIII IIII IIII -
- Cratopus triplicatus IIII v

Dotaz:

- 510- 3/2- 500 1L - Cratopus curvatus
- 1P - Protanabus, C. hirsutus v
- 510- 3/2- 500 4.2 - P. ruficornis v Instrukce 1. post. 1. 1. 1.
- 510- 3/2- 500 5P - C. gr. hirsutus v
- 510- 2/3- 500 1P - C. hirsutus v
- 510- 0/2- 500 (5P) - C. hirsutus v
- 510- 4/2- 500 (2P) - Cratopus triplicatus v (s. 1. 1. 1.)
- stanice s. C. triplicatus 50- 2/2- 500 (2.2)
- ↓
- ↓
- ↓
- ↓

# 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
- v sloupcích proměnné, kterými objekty charakterizujeme
  - druhy, proměnné prostředí

	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			

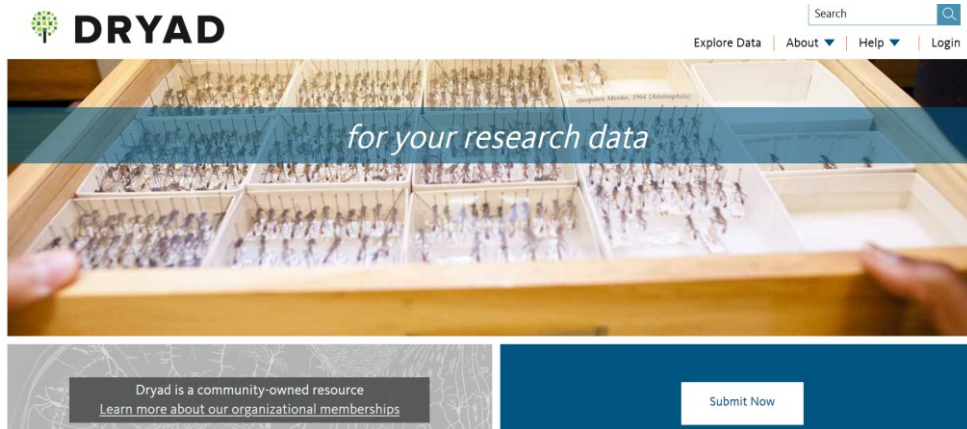
	A	B	C	D	E	F	G	H
1	FIELD_NO	RICHNESS	VEGEALTA	deg_lat	deg_lon	PREC_1	PREC_7	TMEAN_1
2	170-06	95	Forest	51.50142	86.01867	13	74	-14.3
3	171-06	107	Forest	51.53725	85.97119	13	74	-14.3
4	172-06	114	Forest	51.52756	85.96128	13	74	-14.3
5	JD201	43	Open	50.47675	87.62292	9	69	-19.6
6	JD202	18	Forest	50.47717	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
- chybějící data (*missing data, NA*)
  - možnosti jejich nahrazení (interpolace, model)
  - vyloučení proměnné nebo vzorku který má hodně chybějících hodnot
- odlehlé body (*outliers*)
  - EDA – *exploratory data analysis*
- další úpravy:
  - sjednocení taxonomické nomenklatury
  - taxonomická adjustace
  - někdy i vyloučení vzácných druhů (odstranění šumu v datech)

	A	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

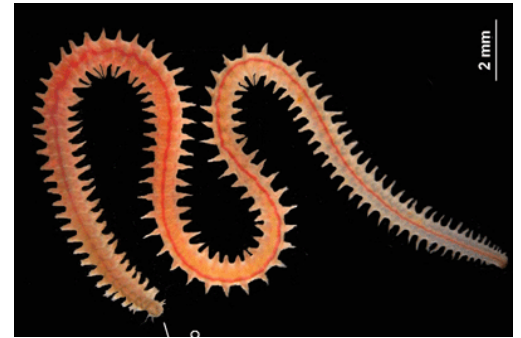


# EXPLORATORY DATA ANALYSIS (EDA)

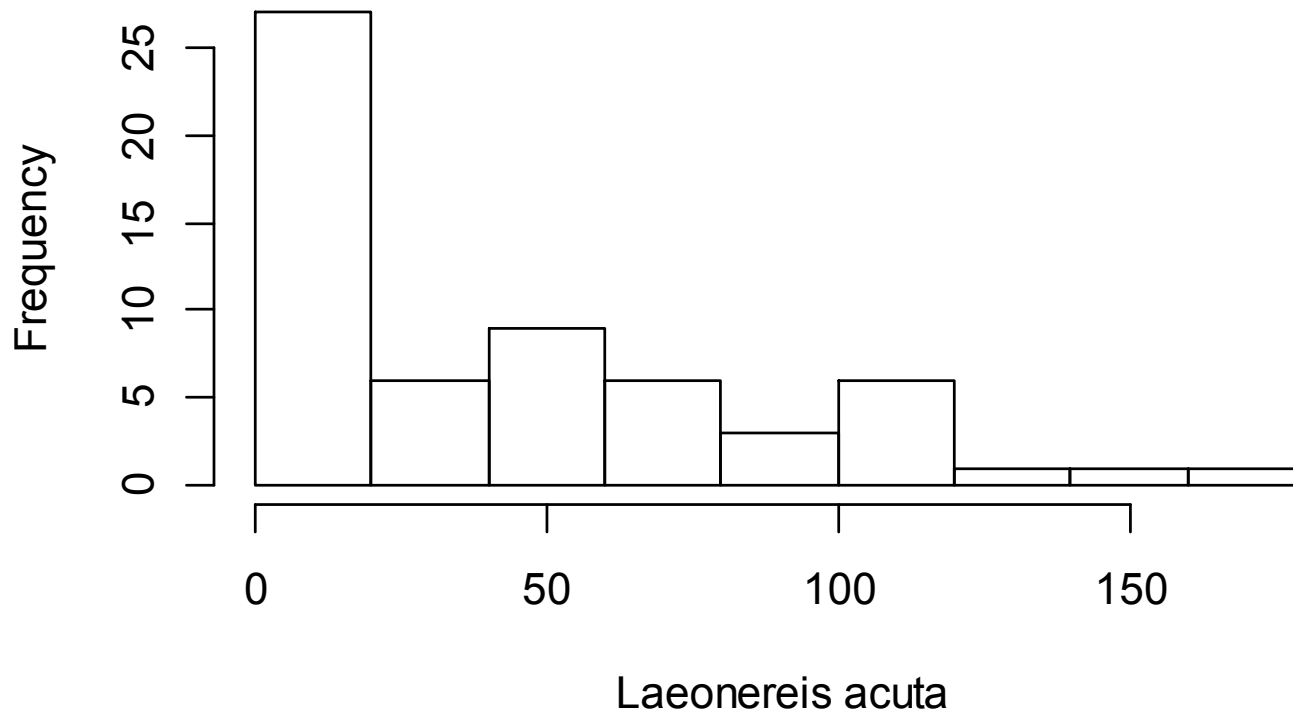
## Proč EDA:

- 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

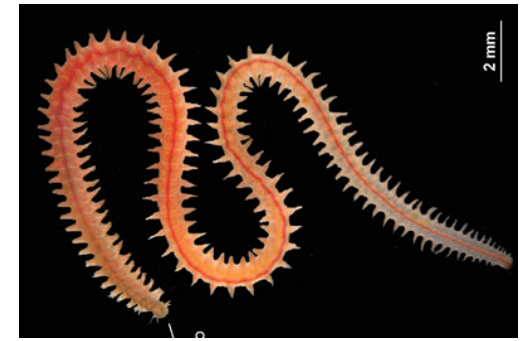
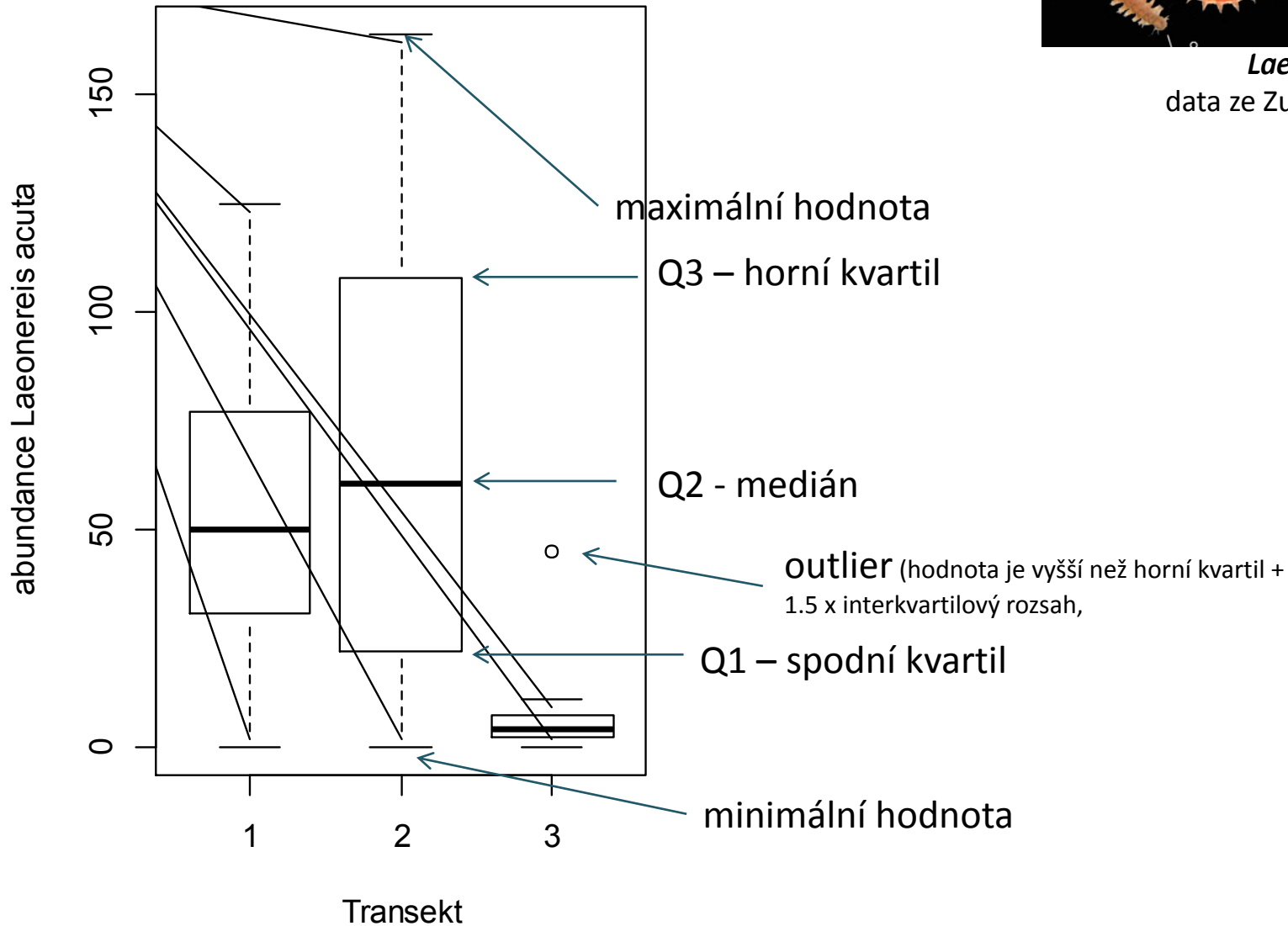
# HISTOGRAM



*Laeonereis acuta*



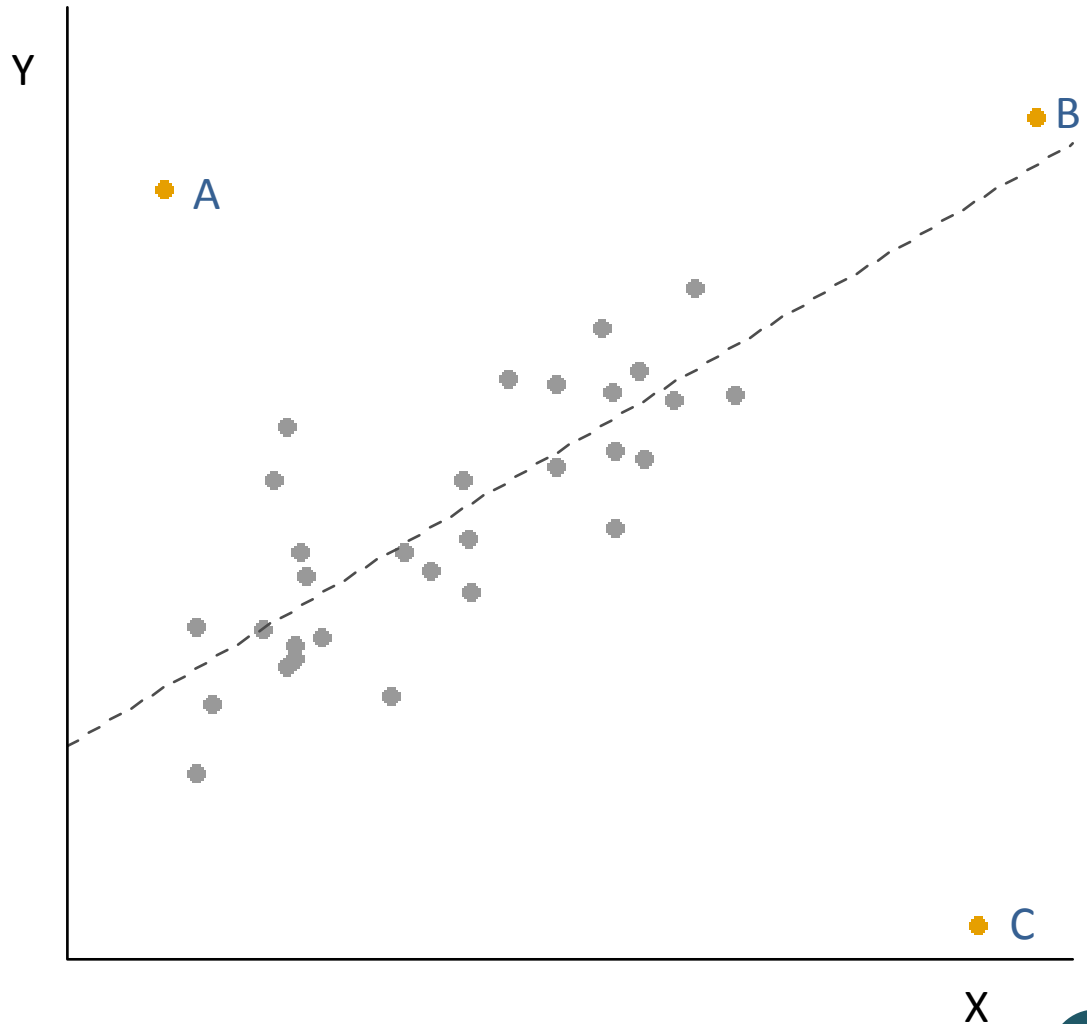
# KONDITIONÁLNÍ BOXPLOT



*Laeonereis acuta*  
data ze Zuur et al. 2007

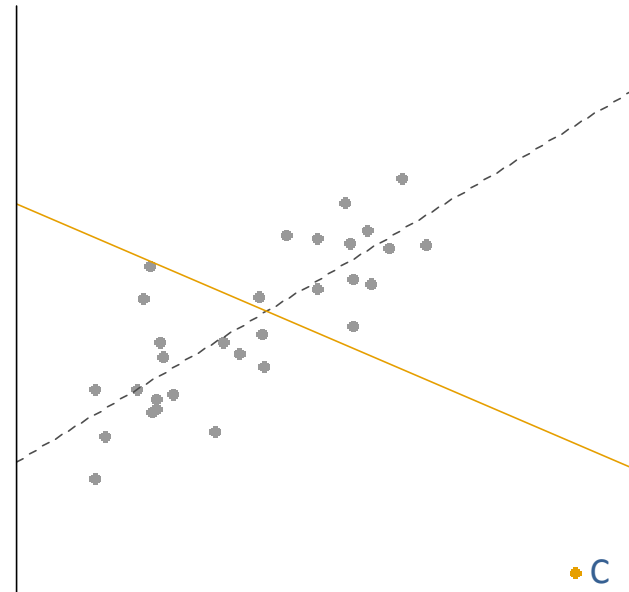
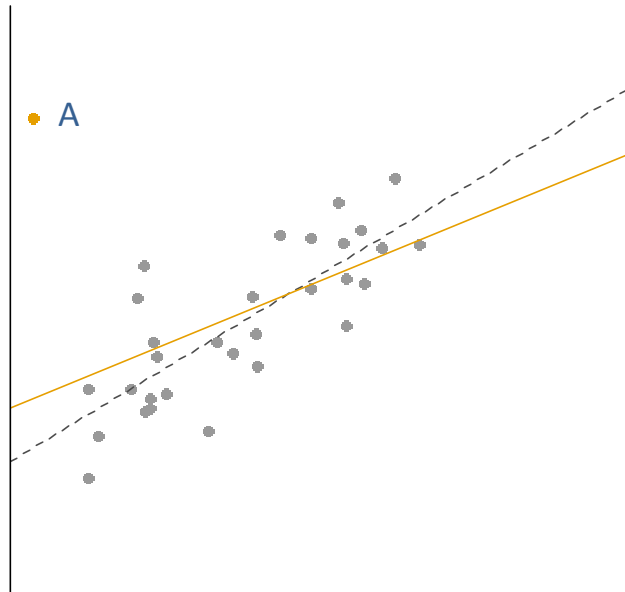
# BODOVÝ GRAF (SCATTER PLOT)

- 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

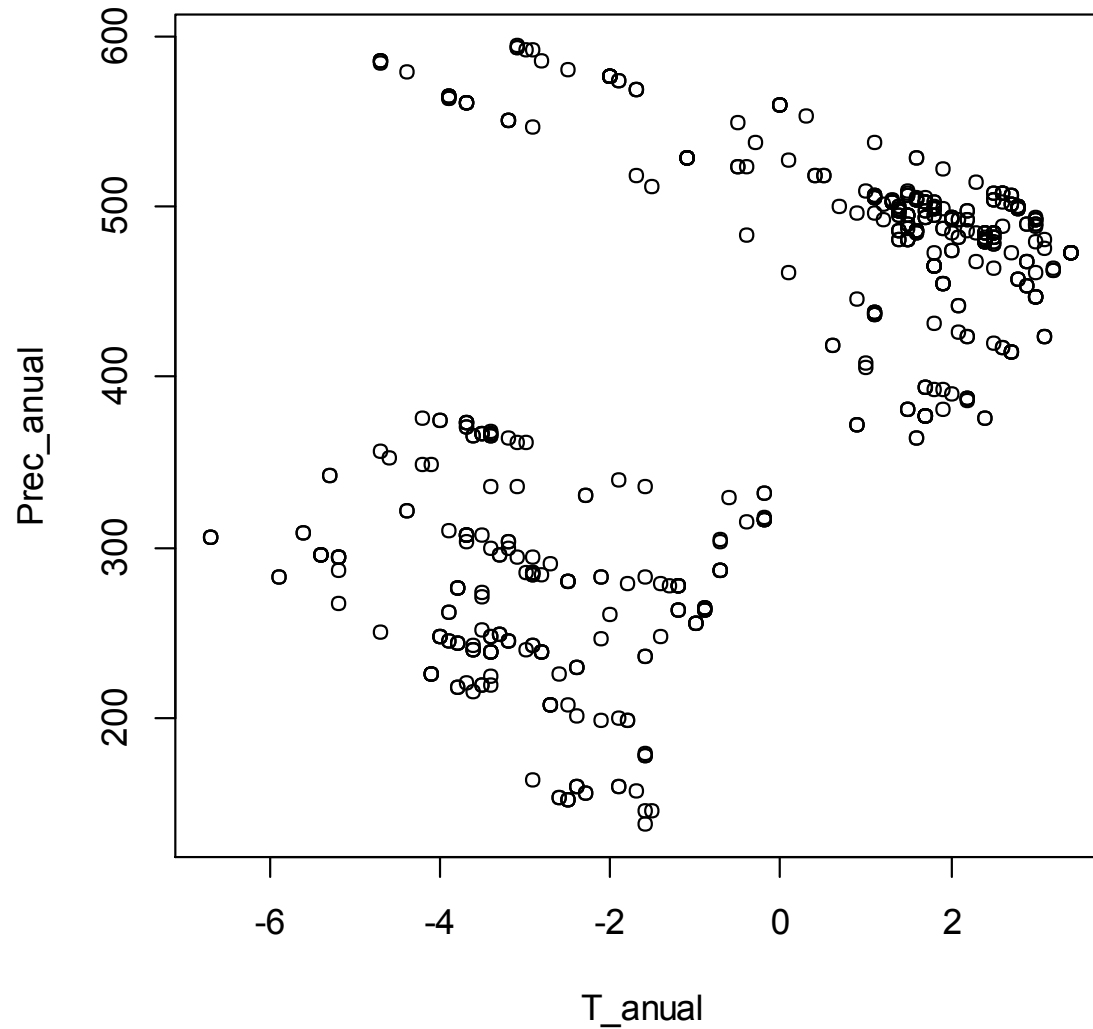


# BODOVÝ GRAF (SCATTER PLOT)

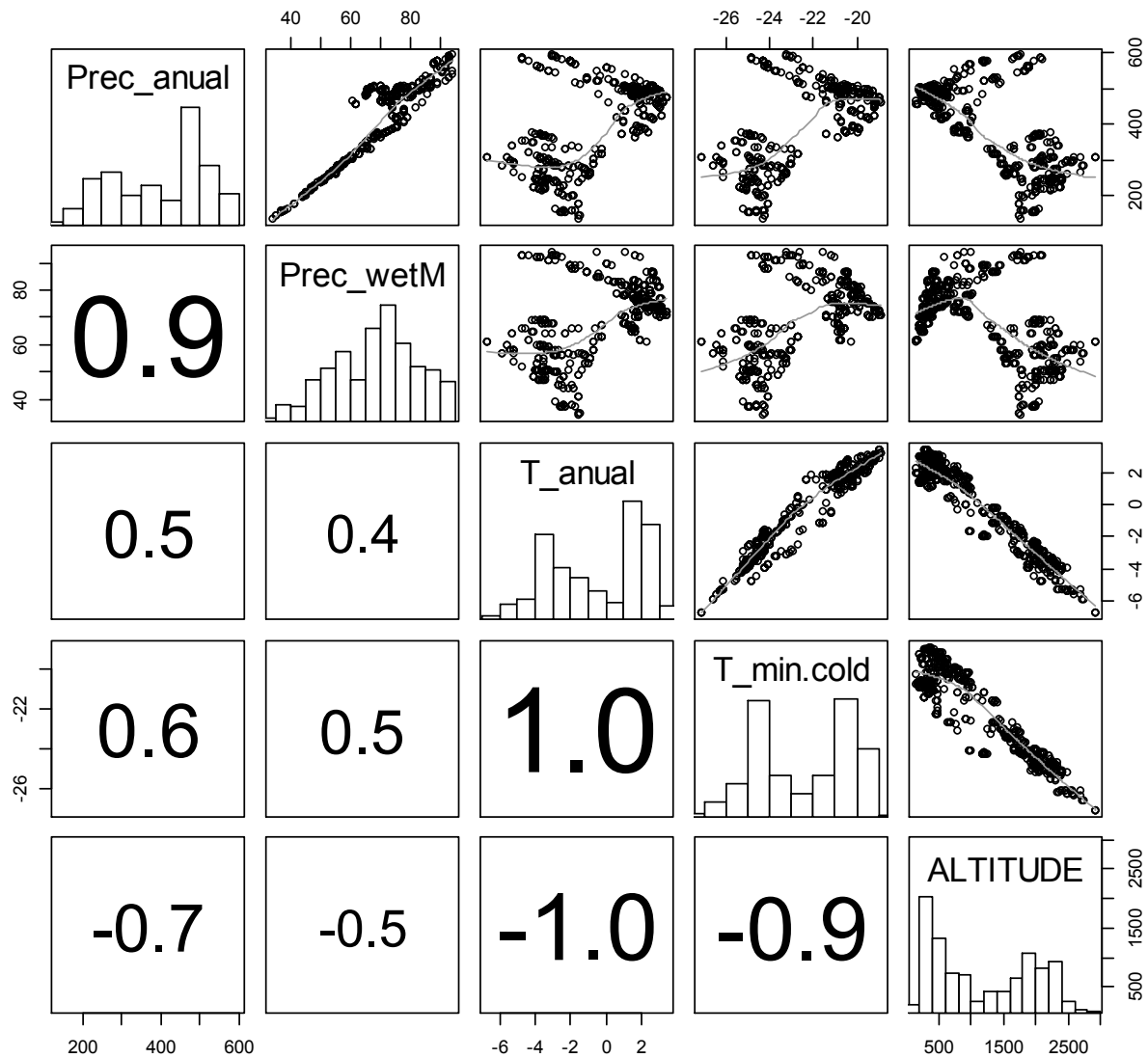
- dramatický vliv outlierů na výsledky analýz



# 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