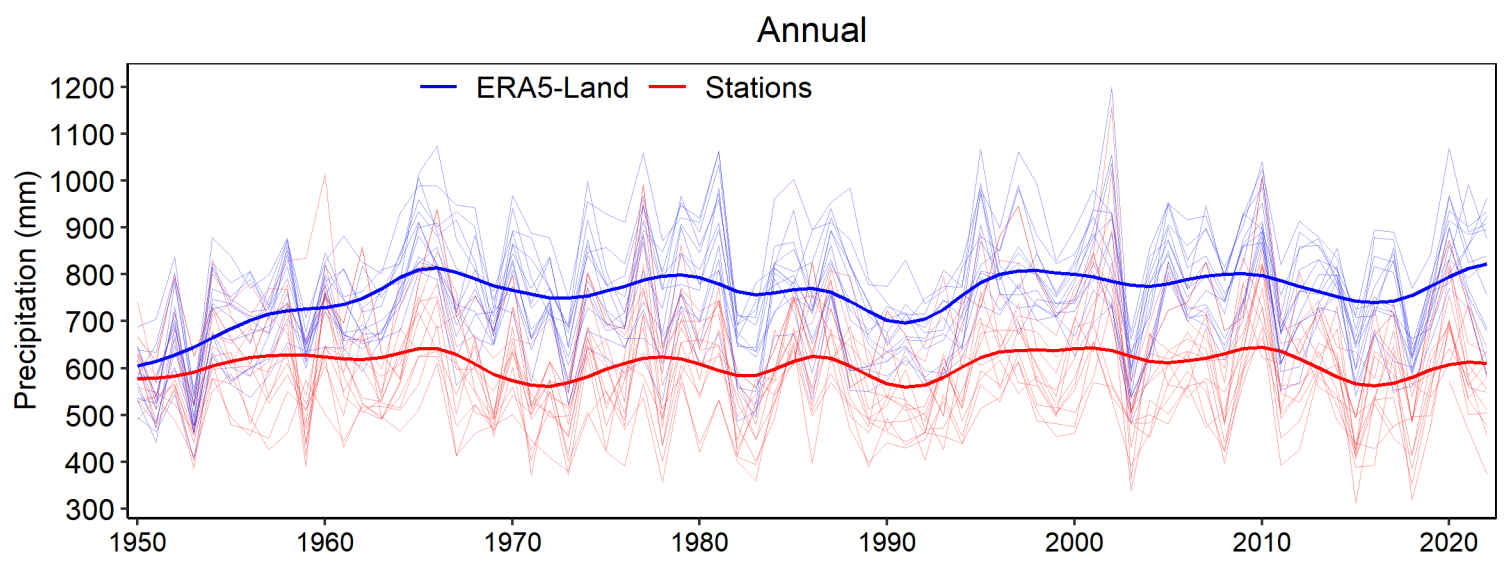
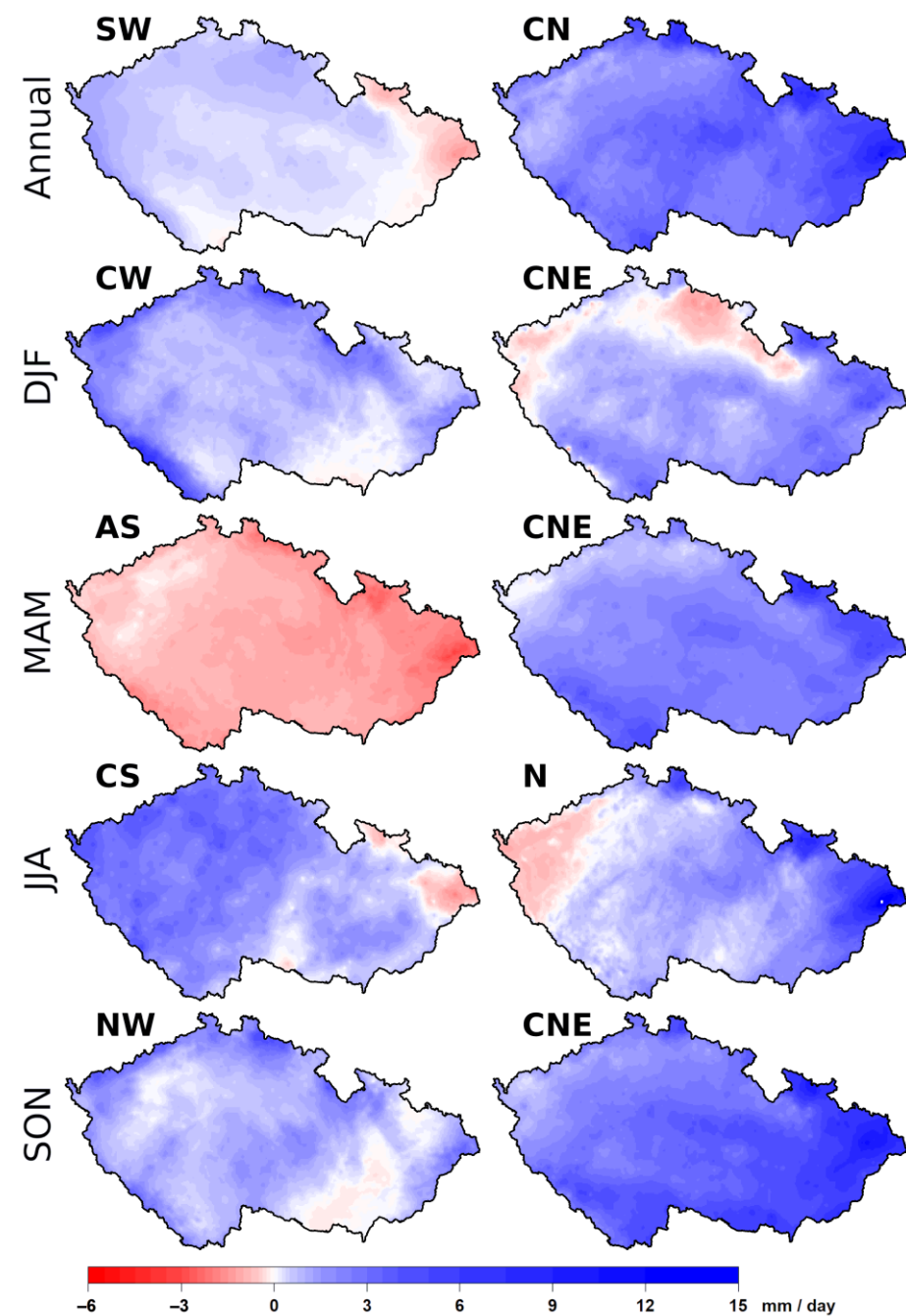
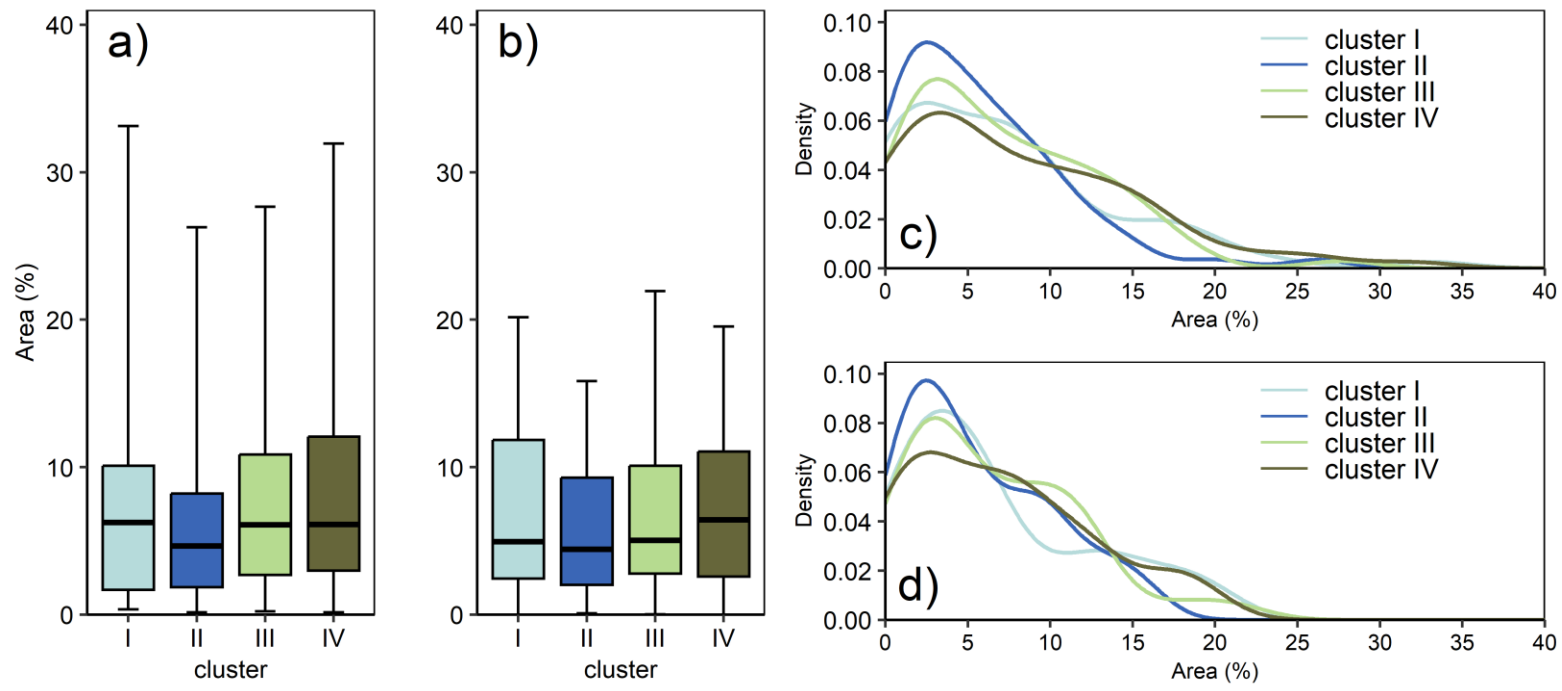


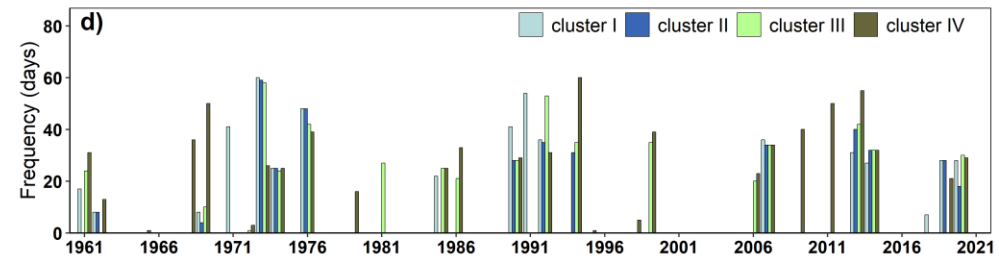
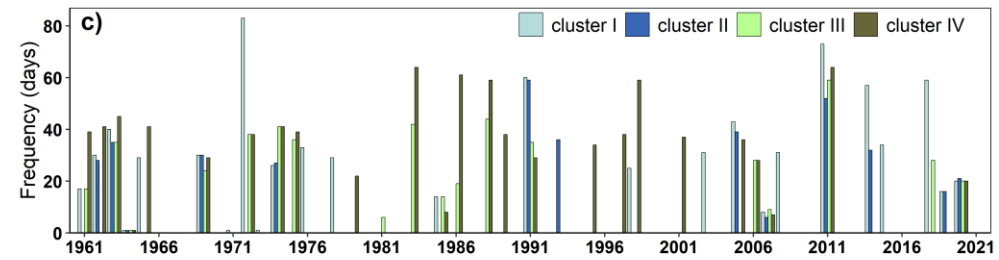
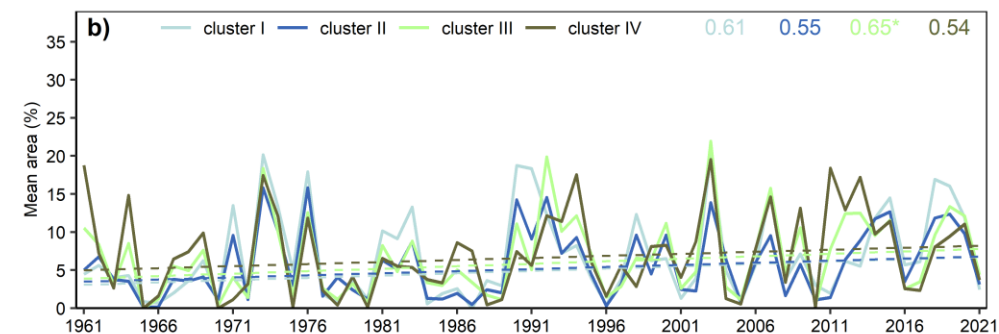
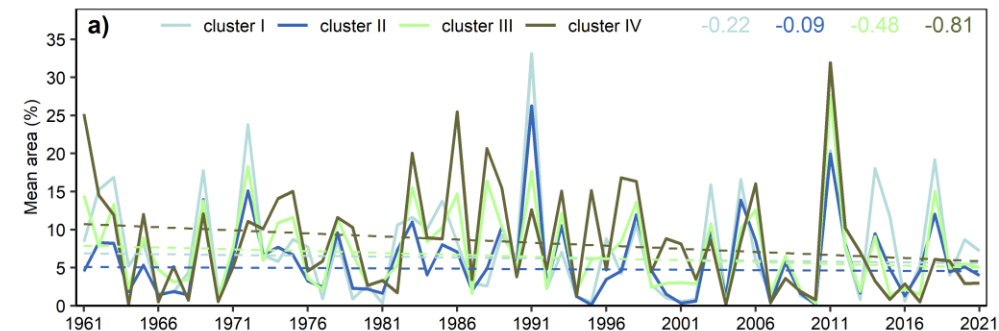
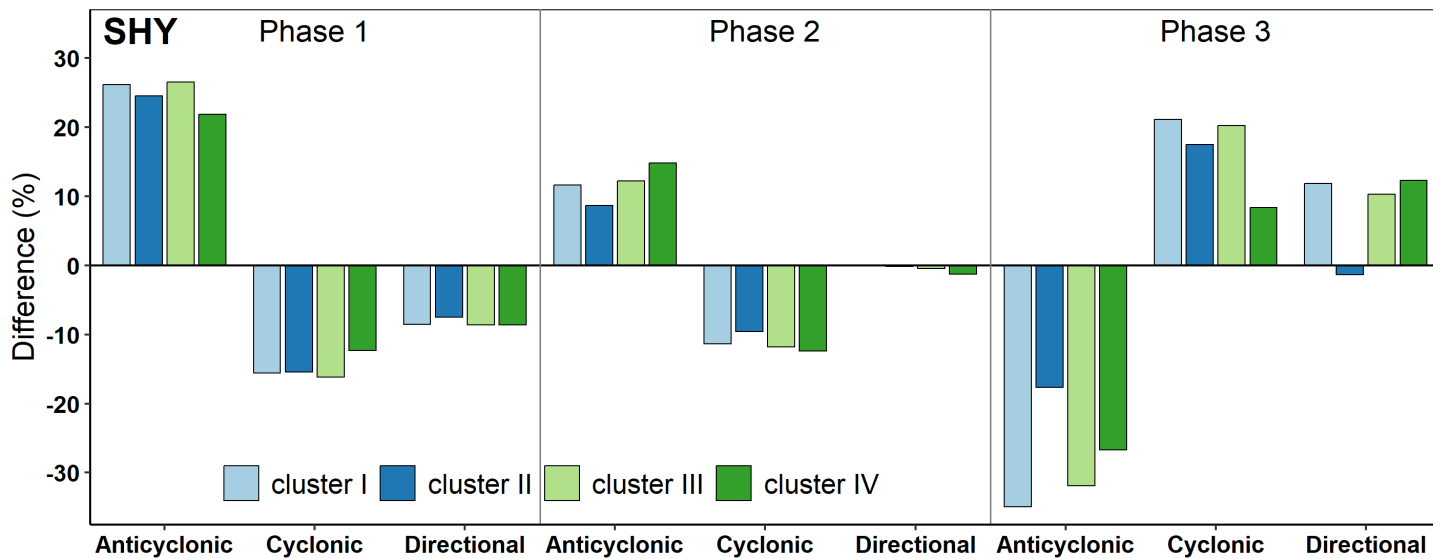
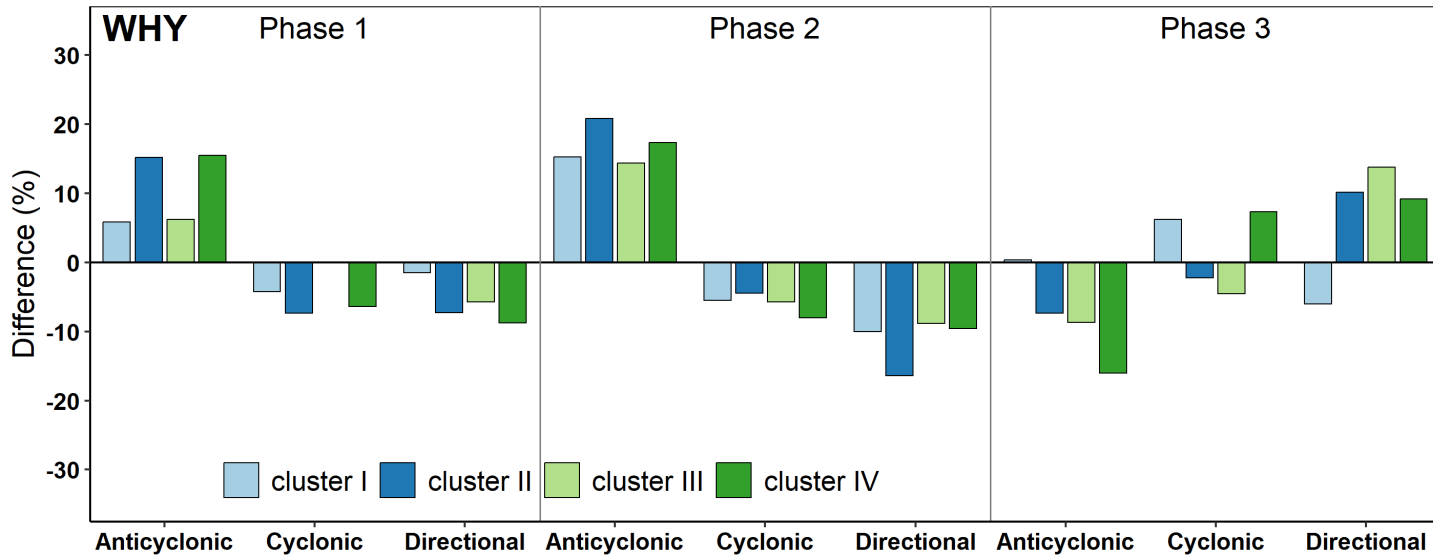
Zpracování a vizualizace dat v R

- **R je programovací jazyk**, jehož prostředí („R-environment“) je určené primárně pro *statistické výpočty, analýzu a vizualizaci dat*
- Je **open-source** => zdarma dostupný, s rozsáhlou komunitou uživatelů a vývojářů
- Používá se například pro:
 - **Statistickou analýzu** (deskriptivní statistika, testy, časové řady)
 - **Práci s daty** => *zpracování velkých objemů tabulárních i prostorových dat*, paralelizace, automatizace
 - **Strojové učení** a „data-mining“
 - **Vizualizaci** – tvorba kvalitních grafů, případně i map (generování velkého množství grafik)
- Klíčová výhoda: **Obrovské množství rozšíření** (Packages = „balíčky/knihovny“)

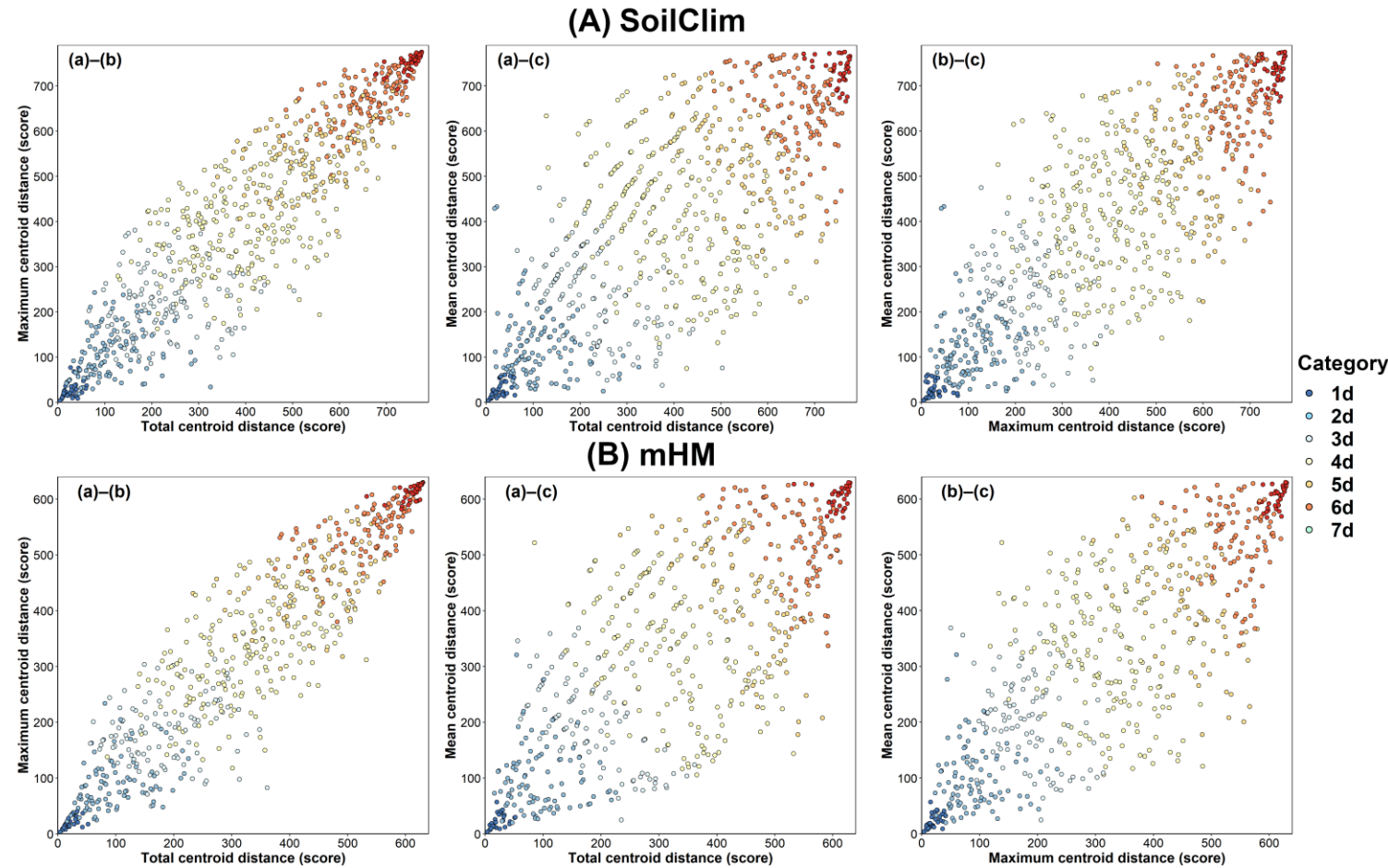
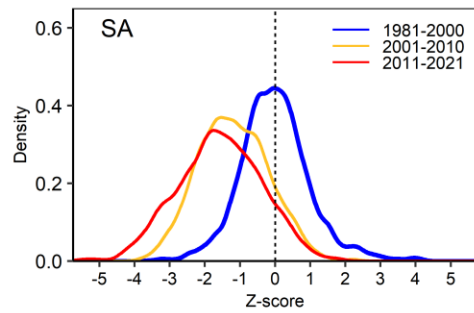
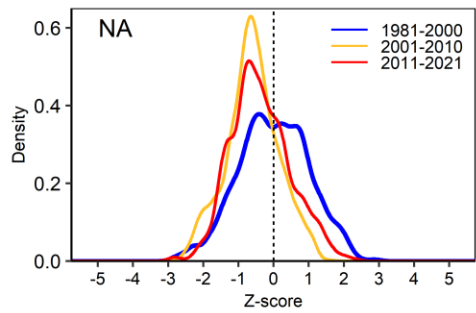
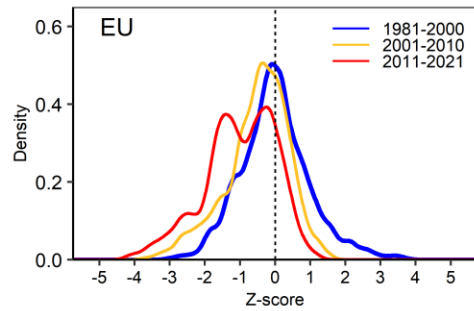
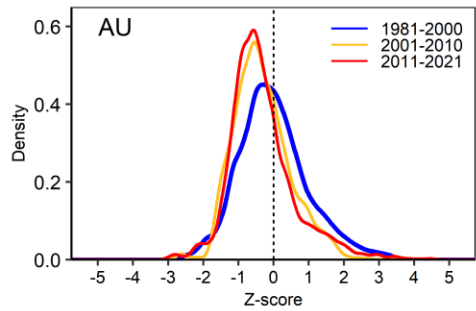
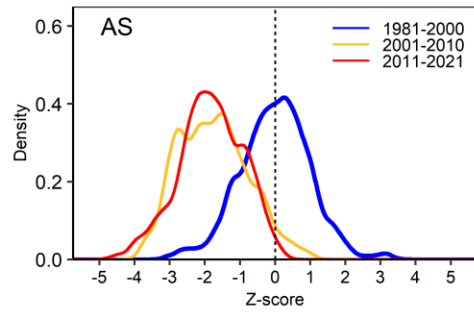
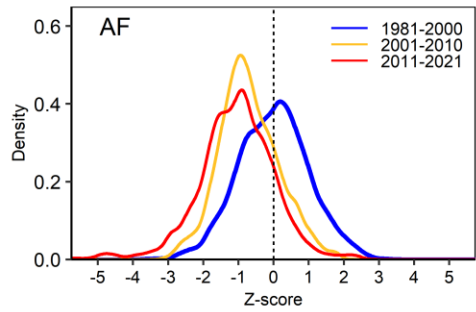
Příklady grafik vygenerovaných přímo R skriptem



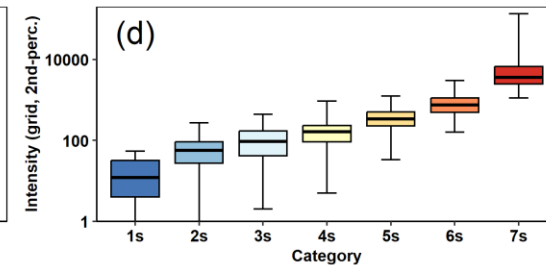
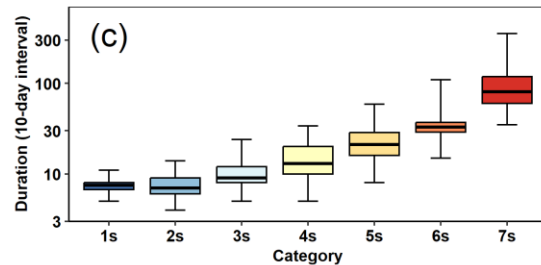
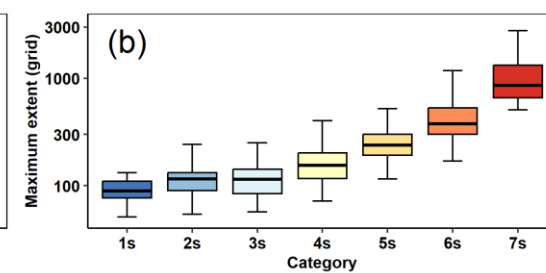
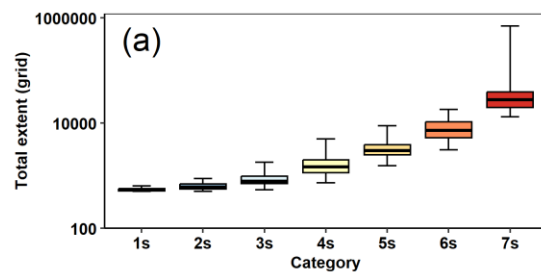
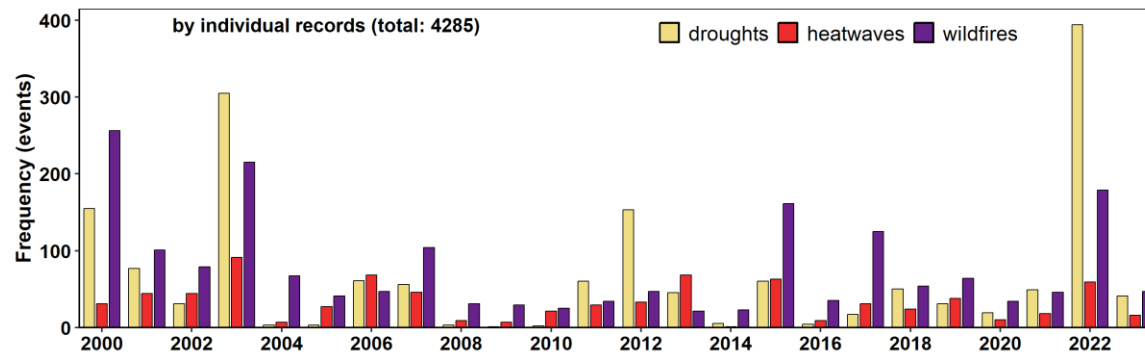
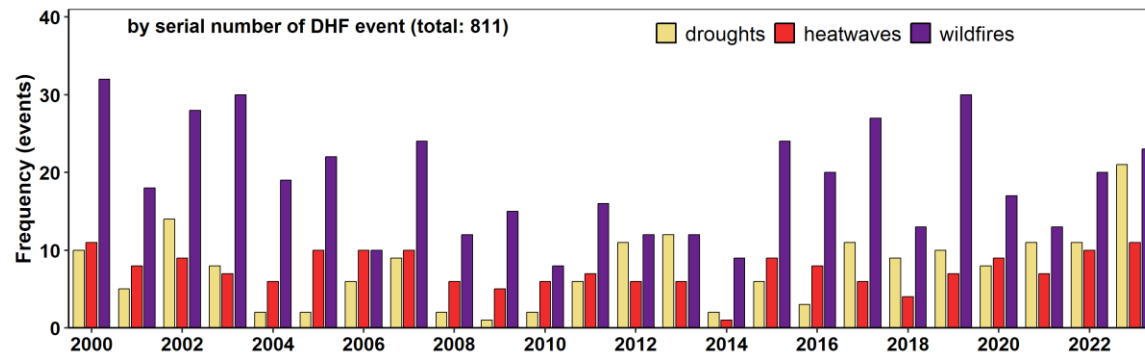
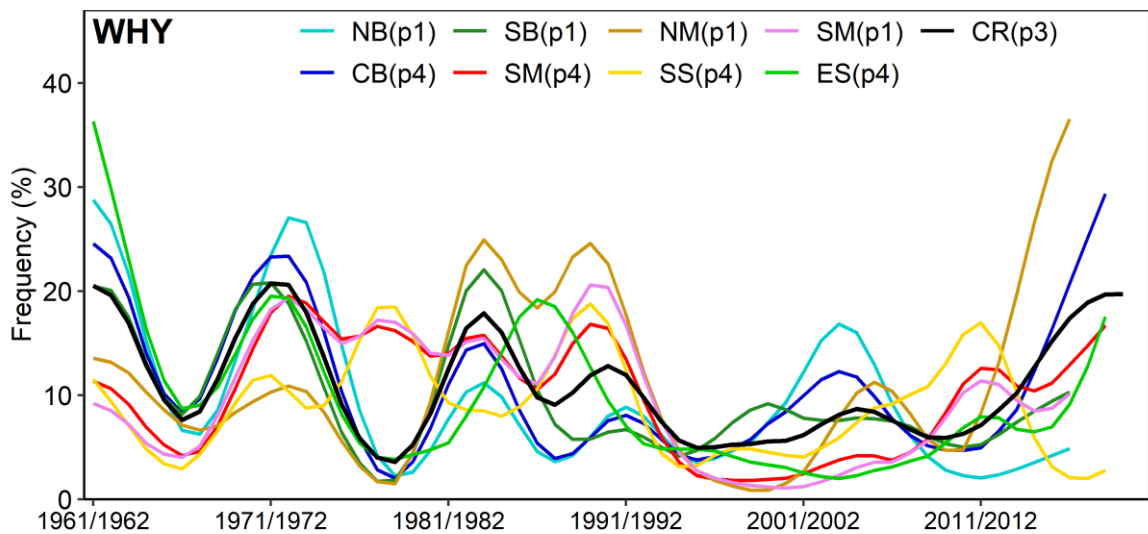
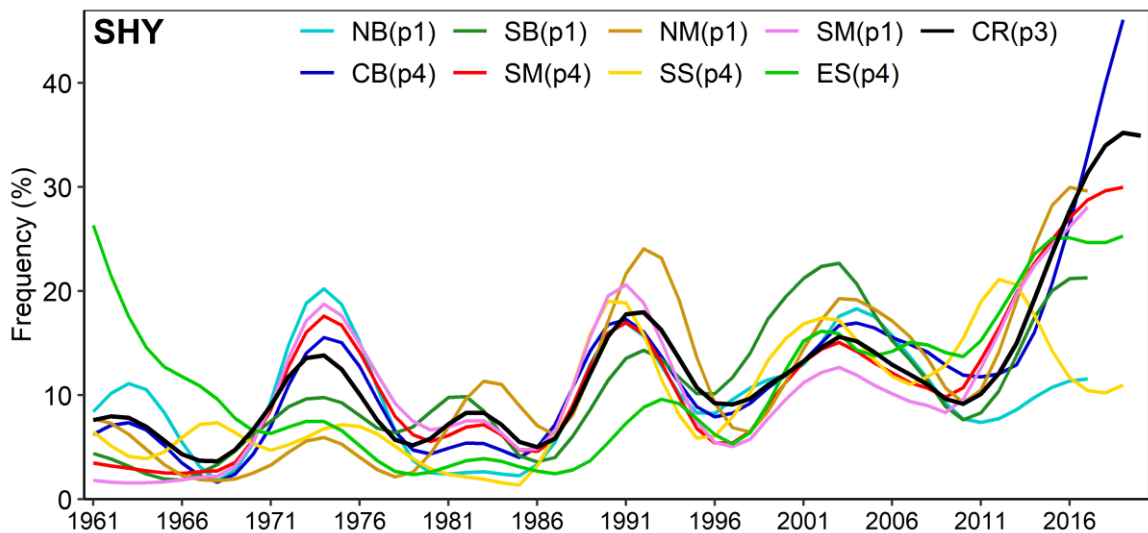
Příklady grafik vygenerovaných přímo R skriptem



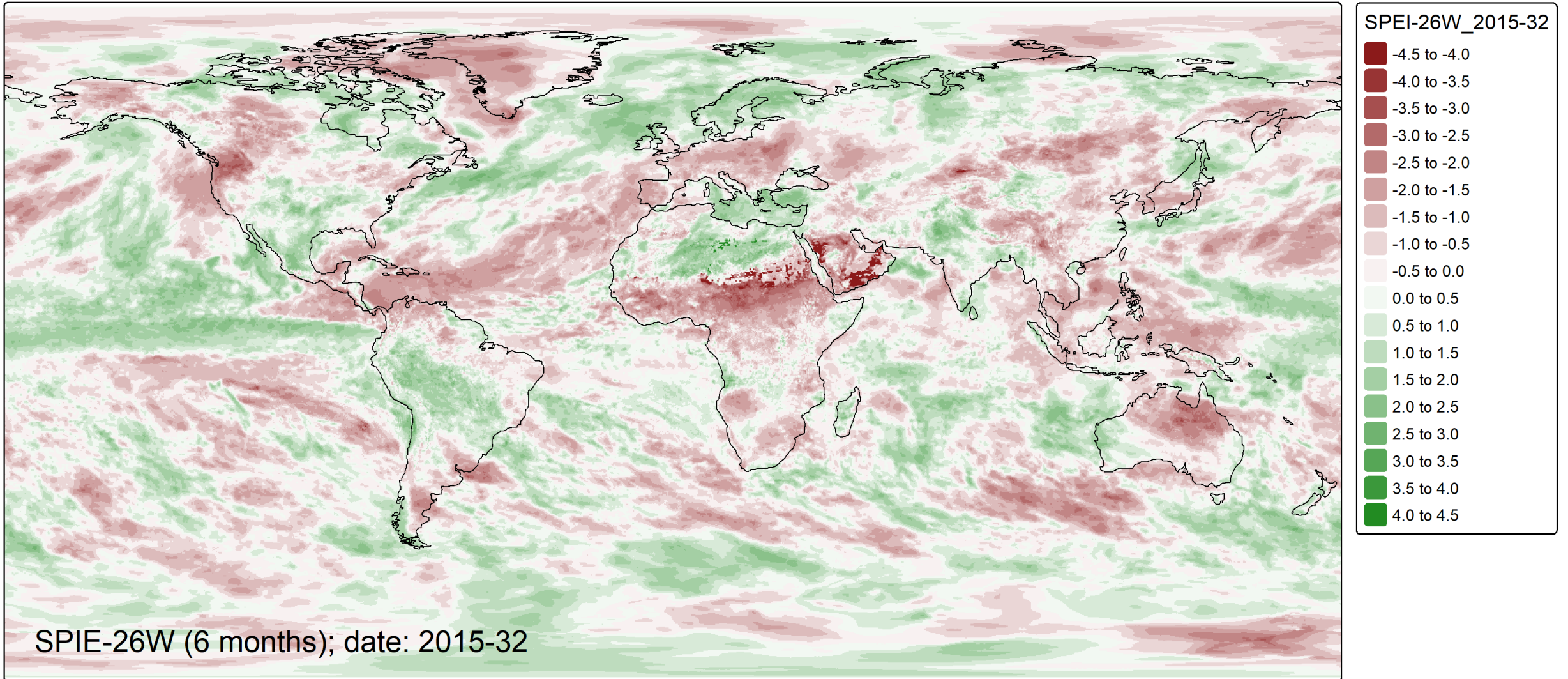
Příklady grafik vygenerovaných přímo R skriptem



Příklady grafik vygenerovaných přímo R skriptem



Příklady grafik vygenerovaných přímo R skriptem



RStudio

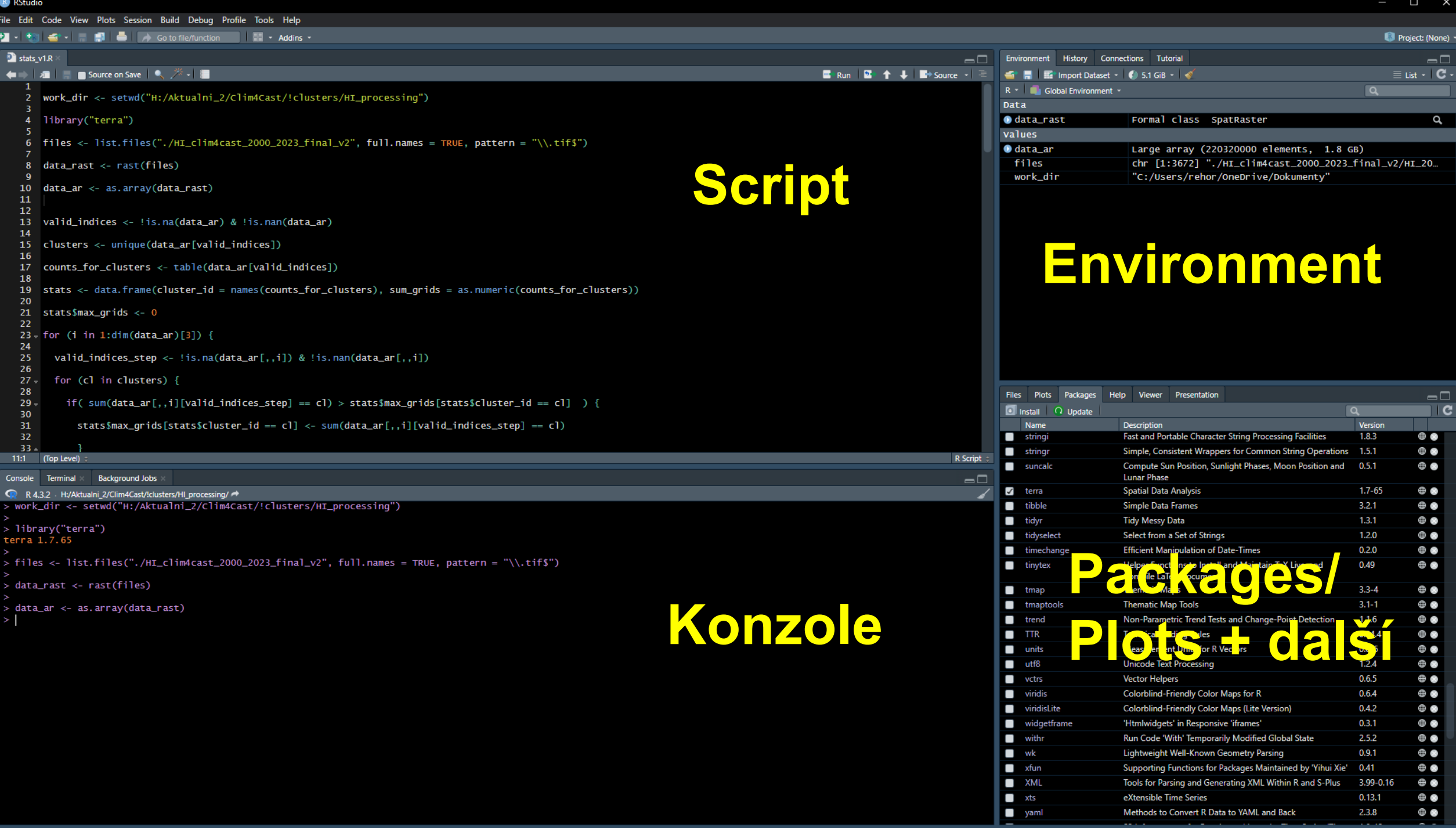
- **Vývojové prostředí** pro R, má několik základních komponent:
- **Console** (konzole) => místo pro **okamžité vykonávání příkazů v R**. Vše, co zde napíšete, se ihned spustí.
- **Script** => okno pro **psaní a ukládání R kódu**. Skripty umožňují spouštět kód po částech nebo celý najednou.
- **Environment** => panel **zobrazující proměnné a další objekty** aktuálně dostupné v paměti.
- **Plots** => sekce, kde se zobrazují vytvořené grafy.
- **Packages** => Knihovny (balíčky) v R jsou **rozšíření jazyka**. Obsahují **funkce**, dataset nebo nástroje pro specifické úlohy. Základní instalace R obsahuje jen **omezenou sadu funkcí**, ale knihovny umožňují rozšířit jeho možnosti.

Script

Environment

Konzole

Packages/ Plots + další



Proměnné v R

- Slouží k *uchování hodnot a objektů* v paměti
- **Přiřazení** se provádí pomocí `<-` nebo `=` (např. `x <- 10`)
- Jsou case-sensitive (X a x jsou různé proměnné)

Základní datové typy

- **Numerické:** čísla (10, 3.14)
- **Celá čísla** (integer): 10L (písmeno "L" označuje celé číslo)
- **Logické** (boolean): TRUE, FALSE
- **Znaky** (character): "text", "Dobré ráno!"
- **Faktory:** speciální typ „character“ dat – kategoriální data ("city", "state") => mají definované tzv. „levels“, tedy všechny hodnoty, které mohou nabývat

Hlavní datové struktury

- **Vektor**, `c()` => základní **jednorozměrná** struktura, např.: `c(1, 2, 3)`; `c("a", "text")`
- **Matice**, `matrix()` => **dvourozměrná struktura** (`matrix(1:9, nrow = 3)`), **tabulková data bez názvů sloupců, vše stejný datový typ**
- **Array (pole)**, `array()` => vícerozměrná matice, typicky 3D
- **Dataframe** (datový „rámeček“), `data.frame()` => **tabulková data, kde každý sloupec je „variable“, má název a může být různý datový typ**
- **Seznam**, `list()` => umožňuje uchovávat různé typy objektů, např.: `list(name="John", age=25, scores=c(90,85,88))`
- Speciální datové struktury pro prostorová data – v rámci knihoven, např. „Spatraster“ nebo „Spatvector“ v rámci knihovny „terra“

Indexování a vektorové operace pomocí []

• Vektor

- Jednotlivý prvek: `x[2]` (druhý prvek vektoru `x`)
- Více prvků: `x[c(1, 3, 5)]` (první, třetí a pátý prvek)
- Vynechání prvků: `x[-2]` (vše kromě druhého prvku)
- Podmíněné filtrování: `x[x > 10]` (prvky větší než 10)

• Matice

- Jednotlivý prvek: `m[2, 3]` (druhý řádek, třetí sloupec)
- Celý řádek/sloupec: `m[2,]` (druhý řádek), `m[, 3]` (třetí sloupec)

• data.frame

- Jednotlivá hodnota: `df[2, 3]` (druhý řádek, třetí sloupec)
- Celý sloupec: `df$sloupec` nebo `df[, "sloupec"]`
- Filtrování podle podmínky: `df[df$vyska > 30,]` (řádky, kde ve sloupci „vyska“ je hodnota > 30)

Funkce

- Základní funkce definované v R (nebo v knihovnách): `mean()`, `sum()`, ...
- Definice vlastní funkce (opakované použití kódu):
 - `my_function <- function(x, y) { return(x + y)}`
 - Volání funkce: `my_function(3, 5) # Výstup: 8`

Základní řídicí struktury v R

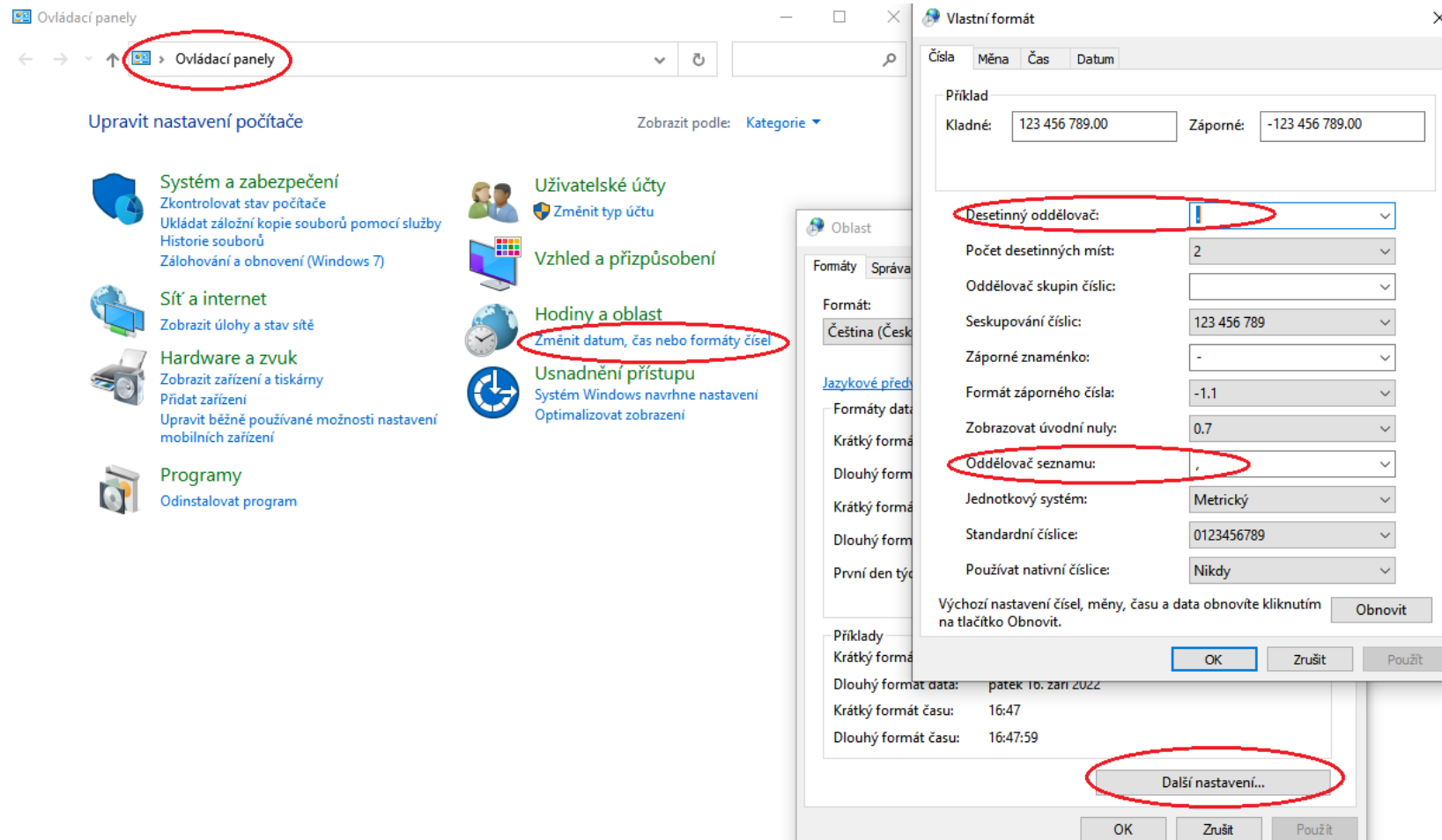
- **Podmínky:** `if (podmínka) { ... } else { ... }`
- **Cykly:**
 - `for (i in 1:10) { ... }` – iterace přes sekvenci
 - `while (podmínka) { ... }` – opakování, dokud platí podmínka
 - `repeat { ... }` – nekonečný cyklus, ukončený pomocí příkazu `break`

Nahrání dat do R

- Základ pro tabulární data je „**csv**“, funkce `read.csv()`
- jiné tabulky `read.table()`, dále spousta dalších možností + další v knihovnách

Anglický formát .csv souborů používaný v R

- Soubory .csv, oddělovač sloupců čárka, desetinná tečka
- Otevření v R => nebo ***read.csv()*** nebo ***read.table(header = TRUE, sep = ",")***
- Otevření v MS Excel => je třeba mít správně nastavené oddělovače ve Windows:



Rozsáhlý výukový materiál k R:

<https://aved.econ.muni.cz/>