

Import a struktury

Cíle

- Datové struktury: vektor a data.frame
- Načtení RDS souborů
- Načtení CSV souborů
- Načtení XLS souborů
- ... ta # \$ @ & % * ! čeština!

Vektor

- Sada hodnot stejného typu
- Funkce se uplatní na všechny prvky vektoru
- Založíme funkcí `c()`: `cisla ← c(1, 2, 3)`
- Konkrétní hodnotu vektoru voláme operátorem hranatá závorka `[]` (česky: pravý Alt + G / F)

Data.frame

- Tabulka souvisejících hodnot (řádky a sloupce)
- Sloupce (proměnné) vždy stejného typu, řádky (pozorování) se mohou v typu lišit
- Sloupec jako vektor voláme operátorem dolar `$` (česky: pravý Alt + ů)

Data.frame jako tabulka

country	year	cases	population
Afghanistan	1999	7745	15557071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	1280428583

variables

country	year	cases	population
Afghanistan	1999	7745	15557071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	1280428583

observations

Soubory typu RDS

- Obsahují 1 seriový objekt, libovolného typu
- Načteme ze souboru funkcí `readRDS()`
- Uložíme na disk funkcí `saveRDS()`
- Oproti jiným formátům (zejména CSV)
významná úspora místa na disku

Soubory typu CSV

- Obecný formát přenosu dat napříč systémy
- Načteme ze souboru funkcí `read_csv()`, respektive `read_csv2()` (české oddělovače)
- Na disk uložíme `write_csv()`, respektive `write_csv2()`

Soubory typu XLS & XLSX

- Starý a nový formát MS Excel
- Načteme ze souboru funkcí `read_excel()`
- Možno omezit na určitou oblast dat
- Pro přenos dat do Excelu použijte CSV :)

Kódování češtiny

- Při ručním importu:
 - Delimeter & Locale / Decimal Mark
 - Locale / Encoding
- Když už se rozsypaný čaj dostane do erka:
 - `library(stringi)`
 - `stri_encode(str, from, to)`