

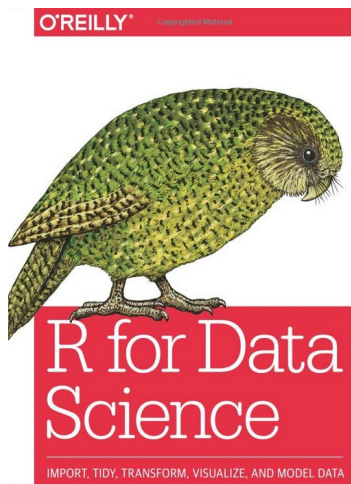
Úvod do R – Manipulace dataframu

Vít Syrovátka

Tidyverse

- soubor balíčků se stejnou filozofií
- cílem je zpřehlednění a zjednodušení kódu
- výhoda – přehledný kód bez mnoha vnořených závorek a bez uvozovek
- nevýhoda – nepodporuje jména řádků!

Hadley Wickham



Hadley Wickham &
Garrett Grolemund

Copyrighted Material

Tidyverse



magrittr pipes %>%

- použije výsledek předchozí operace jako první argument následující funkce
- přehlednější psaní kódu
- kód je zapisován v takové posloupnosti, v jaké uvažujeme
- zapadá do koncepce `tidyverse`

příklad: jaká je průměrná délka nohy mužů, zaokrouhlaná na celá čísla?

```
> round(mean(miry$noha[miry$sex == "m"]))
```

magrittr pipes %>%

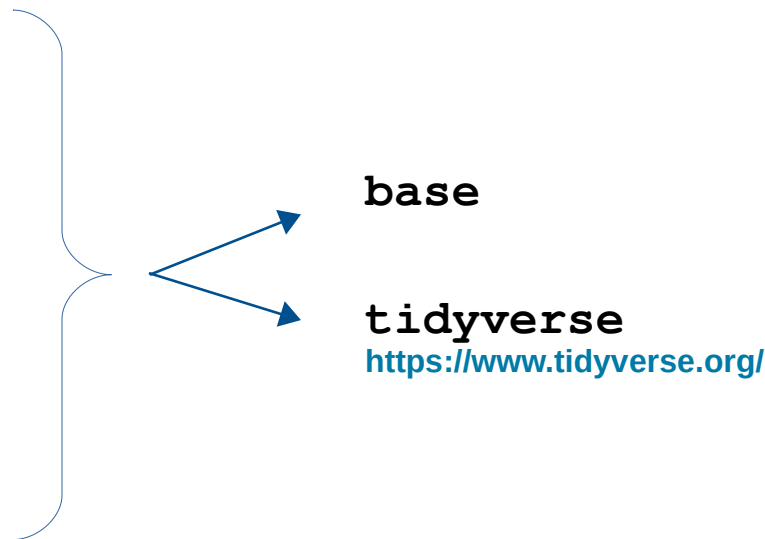
- použije výsledek předchozí operace jako první argument následující funkce
- přehlednější psaní kódu
- kód je zapisován v takové posloupnosti, v jaké uvažujeme
- zapadá do koncepce tidyverse

příklad: jaká je průměrná délka nohy mužů, zaokrouhlaná na celá čísla?

```
> miry %>%  
>   select(noha) %>%  
>   filter(sex == "m") %>%  
>   mean %>%  
>   round
```

Časté manipulace

- vytvoření nové proměnné
- úprava stávající proměnné
- odstranění proměnné
- výběr proměnných
- výběr pozorování (cases)



Vytvoření, úprava proměnné

- `jmeno_dataframu$jmeno_promenne <- data`

```
> miry$vyska_m <- miry$vyska / 100
```

- `dplyr::mutate`

```
> miry <-  
>   miry %>%  
>   mutate(vyska_m = vyska / 100)
```

Odstranění proměnné

- `jmeno_dataframu$jmeno_promenne <- NULL`

```
> miry$vyska <- NULL
```

- `dplyr::mutate`

```
> miry <-  
>   miry %>%  
>   mutate(vyska = NULL)
```

- nebo vybereme jen ty proměnné, které chceme zachovat...

Výběr proměnných

- `jmeno_dtf[numericke_subscript]`

`jmeno_dtf[jmenny_subscript]`

`jmeno_dtf[logicky_subscript]`

- `dplyr::select`

```
> miry %>%  
>   select(1:2)  
  
> miry %>%  
>   select(sex, noha)
```


Výběr pozorování

- `jmeno_dtf[numericke_subscript,]`
`jmeno_dtf[jmenny_subscript,]`
`jmeno_dtf[logicky_subscript,]`
- `dplyr::filter`

```
> miry %>%  
>   dplyr::filter(sex == "z")  
  
> miry %>%  
>   dplyr::filter(sex == "m" | vyska > 175)
```