

# Portál:R wiki (pro pokročilejší)



Portál **R wiki** je věnován programovacímu jazyku **R** (<http://www.r-project.org/>) a jeho aplikacím v biomedicině. Důraz je na jedné straně kladen na jednoduché a postupné seznamování s jazykem, předpokládající jeho nulovou původní znalost, na druhé straně však i pokus o publikování pokročilých řešení netriviální problematiky za hranou dostupné anglofonní dokumentace.

## Užitečné odkazy

### České zdroje

- Přehled materiálů a zdrojů (<http://www.karlin.mff.cuni.cz/~kulich/vyuka/Rdoc/index.html>) od doc. Kulicha, včetně jeho Stručného úvodu do R (<http://www.karlin.mff.cuni.cz/~kulich/vyuka/Rdoc/uvodrfpm.pdf>)
- Základy práce s R (<http://www.karlin.mff.cuni.cz/~komarek/vyuka/dataRko/Rmanual1.pdf>) a Hrátky s R ([http://www.karlin.mff.cuni.cz/~kulich/vyuka/Rdoc/arnost\\_introR.pdf](http://www.karlin.mff.cuni.cz/~kulich/vyuka/Rdoc/arnost_introR.pdf)), oboje od doc. Komárka
- Jak na jazyk R (<http://www.karlin.mff.cuni.cz/~komarek/vyuka/dataRko/Rmanual2.pdf>) od doc. Bíny, doc. Komárka, doc. Komárkové
- Cvičení z biostatistiky (<http://cran.r-project.org/doc/contrib/CviceniR1.pdf>) v R, od doc. Drozda
- Jak pracovat s jazykem R ([http://www.math.muni.cz/~xkonecn3/vyuka/MUNI/VMS/navod\\_R.pdf](http://www.math.muni.cz/~xkonecn3/vyuka/MUNI/VMS/navod_R.pdf)), od mgr. Konečné a doc. Kolářka
- ZVÁRA, Karel. *Základy statistiky v prostředí R*. 1. vydání. Praha : Karolinum, 2013. Biomedicínská statistika; sv. 4. ISBN 978-80-246-2245-3.

### Zahraniční zdroje

- **Domovská stránka jazyka R** (<http://www.r-project.org/>)
- CRAN (<https://cran.r-project.org/>), archiv většiny dostupných rozšiřujících balíčků R
- The R Manuals (<http://cran.r-project.org/manuals.html>), anglicky
  - An Introduction to R (<http://cran.r-project.org/doc/manuals/r-release/R-intro.html>)
- Quick-R (<https://www.statmethods.net/>), přehledný a praktický průvodce jazykem R
- R-bloggers (<https://www.r-bloggers.com/>), novinky a tutoriály ze světa R
- R for cats (<https://rforcats.net/>), manuál R určený kočkám domácím (*Felis silvestris*)
- R for Data Science (<https://r4ds.had.co.nz/>) Volně přístupná učebnice R pro práci s daty
- Advanced R (<http://adv-r.had.co.nz/>), web Hadleyho Wickhama, otce moderního R
- The R Inferno ([http://www.burns-stat.com/pages/Tutor/R\\_inferno.pdf](http://www.burns-stat.com/pages/Tutor/R_inferno.pdf)), o nástrahách R a jak se jim vyhnout
- dotazy ohledně R na Stack Overflow (<https://stackoverflow.com/questions/tagged/r>)
- #rstats (<http://twitter.com/hashtag/rstats>) hashtag na Twitteru (<http://twitter.com/>)

### Online platformy, kurzy

- R-Fiddle (<http://www.r-fiddle.org/#/>), online konzole jazyka R
- RStudio (<https://www.rstudio.com/home/>), open-source vývojové prostředí jazyka R
- Try R! (<https://www.pluralsight.com/search?q=R>), otevřená online škola kódování v R, **vhodná i pro začátečníky**
- DataCamp (<https://www.datacamp.com/courses>), online dostupné kurzy jazyka R
- Coursera (<https://www.coursera.org/browse/data-science>), volně dostupné kurzy týkající se (nejen) statistiky, analýzy dat, R

- Statistics with R (<https://www.coursera.org/speci>

### Lidé kolem jazyka R

- Hadley Wickham (<http://adv-r.had.co.nz/>), osobní web otce moderního R
- Yihui Xie (<https://yihui.name/>), osobní web autora několika důležitých R balíčků, především balíčku knitr (<http://cran.r-project.org/web/packages/knitr/index.html>)
- Rob J Hyndman (<https://robjhyndman.com/>), osobní web australského statistika známého prediktivními modely

### Události kolem jazyka R

alizations/statistics) ■ Data Science (<https://www.coursera.org/specializations/jhu-data-science>) ■ Introduction to Neurohacking In R (<https://www.coursera.org/learn/neurohacking>)

- edX (<https://www.edx.org/course/subject/data-analysis-statistics>), volně dostupné kurzy týkající se (nejen) statistiky, analýzy dat, R

■ Data Analysis for Life Sciences (<https://www.edx.org/xseries/data-analysis-life-sciences>) ■ Genomics Data Analysis (<https://www.edx.org/xseries/genomics-data-analysis>) ■ Introduction to R for Data Science (<https://www.edx.org/course/introduction-r-data-science-microsoft-dat204x-3>) ■ Programming with R for Data Science (<https://www.edx.org/course/programming-r-data-science-microsoft-dat209x-2>) ■ Foundations of Data Analysis - Part 1: Statistics Using R (<https://www.edx.org/course/foundations-data-analysis-part-1-utaustinx-ut-7-11x-0>) ■ Explore Statistics with R (<https://www.edx.org/course/explore-statistics-r-kix-kiexplorx-0>)

- Intellipaat (<https://intellipaat.com/>), kurz dostupný online související s

■ Data Scientist course (<https://intellipaat.com/data-scientist-course-training/>) ■ Machine Learning Course (<https://intellipaat.com/machine-learning-certification-training-course/>) ■ Artificial Intelligence Certification (<https://intellipaat.com/artificial-intelligence-deep-learning-course-with-tensorflow/>).

- Statistical Learning (<http://online.stanford.edu/courses>), online dostupný stanfordský kurz strojového učení v R
- swirl (<https://swirlstats.com/>), balíček na procvičování R, spouští se přímo v R

- useR! (<http://www.r-project.org/conferences.html>), každoroční mezinárodní konference určená pro širokou komunitu jazyka R

■ odkaz na ročník konference 2016 na Stanfordské univerzitě (<http://user2016.org/MZYdZ/QPkoZ/nLWYP/>)

- Directions in Statistical Computing (DSC) (<http://www.r-project.org/conferences.html>), mezinárodní konference určená vývojářům statistického software
- Prague R Meetup (<https://www.meetup.com/Prague-R-Meetup-Group/>), občasná pražská setkání **české** komunity kolem jazyka R

## Úvod

- Proč bych se měl právě já učit R
- Co je R
- Srovnání R s dalšími podobnými jazyky

## Začínáme

- Stažení a instalace jádra jazyka R
  - Instalace ve Windows ■ Instalace na Mac ■ Instalace v Linux
- Odvozený a další doporučený software
  - RStudio ■ Textový editor ■ TeX ■ Pandoc
- První spuštění
  - Práce s nápovědou a kde ji získat
  - Přístup k práci v R
    - Interaktivní používání R ■ Používání R na základě volání skriptů
- Instalace nadstavbových balíčků
  - Reinstalace balíčků při aktualizaci jádra R

## Přehled základních datových typů a struktur

- Datové typy

- Numerická hodnota ■ Celé číslo ■ Komplexní hodnota ■ Logická hodnota ■ Řetězec ■ NA ■ NULL ■ NaN
- Přetypování a koerce
- Datové struktury
  - Vektor (R) ■ Faktor ■ Matice ■ Pole ■ Tabulky dat ■ Seznam
- Mody a atributy objektů
  - Vnitřní atributy: mod a délka ■ Získání a změna atributů ■ Třída objektu

## Jednoduché operace, čísla, vektory

- Tvorba vektorů a základní příkazy
  - Tvorba pravidelných sekvencí
- Logické vektory
- Vektory řetězců
- Chybějící hodnoty
- Subvektory, indexování, adresace
- Faktory
  - Seřazené faktory
- Matematické operátory a operace
  - Aritmetické operátory ■ Logické operátory a porovnávání ■ Množinové operátory ■ Matematické funkce ■ Zaokrouhlování, formátování čísel ■ Konstanty

## Matice a pole

- Tvorba matic a základní příkazy
  - Tvorba kontingenční tabulky
- Manipulace s maticemi
  - Sčítání a odčítání matic ■ Násobení matic ■ Inverzní a transponovaná matice ■ Další maticová aritmetika
- Submatice, indexování, adresace
- Tvorba polí a základní příkazy
- Manipulace s polem
- Subpole, indexování, adresace

## Datové tabulky a seznamy

- Tvorba datových tabulek a základní příkazy
- Manipulace s datovými tabulkami
- Indexování, adresace v datových tabulkách
- Tvorba seznamů a základní příkazy
- Manipulace se seznamy
- Subseznam, indexování, adresace

## Načítání dat do R z externích zdrojů a jejich ukládání

- Načítání dat z místních souborů
  - Funkce read.table() ■ Funkce scan() ■ Funkce readLines() ■ Funkce read.csv() ■ Funkce read.xlsx() pro načítání dat z Excelu ■ Načítání dat z Wordu
- Načítání online dat
  - balíček RCurl
- Načítání dat z jiných balíčků R
  - balíček datasets
- Preprocessing načtených dat

- Ukládání výstupů na místní disk
  - Příkaz sink() - sink() ■ Funkce writeLines() ■ Funkce capture.output() ■ Funkce write.csv() ■ Funkce write.xlsx() pro ukládání dat do Excelu ■ Ukládání grafických výstupů na místní disk

## Funkce v R

- Některé užitečné vestavěné funkce
  - Funkce sessionInfo() ■ Funkce str() ■ Funkce dim() ■ Funkce rbind() ■ Funkce cbind() ■ Funkce length() ■ Funkce Sys.info() ■ Funkce Sys.Date() ■ Funkce class() ■ Funkce weekdays()
- Uživatelem definované funkce
  - Jednoduché příklady vlastních funkcí ■ Argumenty funkce ■ Výstupy funkce ■ Přiřazování v těle funkce ■ Rekurse ■ Pokročilejší příklady vlastních funkcí

## R jako programovací jazyk

- Úrovně prostředí, scoping
- Globální a lokální proměnné
- Podmínky a podmíněné příkazy
  - Podmínka když
- Cykly
  - Cyklus for-do ■ Cyklus while-do ■ Cyklus repeat-until
- Varování a chyby
- Skupina příkazů apply()
  - Funkce apply() ■ Funkce lapply() ■ Funkce sapply() ■ Funkce vapply() ■ Funkce tapply() ■ Funkce replicate()
- Pravidla smart code
  - Smart code dle Google ■ Smart code dle Hadleyho Wickhama

## Statistika v R

- Náhodná proměnná, pravděpodobnostní rozdělení
  - Bernoulliho rozdělení (R) ■ Binomické rozdělení (R) ■ Geometrické rozdělení (R) ■ Normální rozdělení (R) ■ Poissonovo rozdělení (R)
- Deskriptivní statistika (R)
  - Míry polohy (R) ■ Míry variability (R) ■ Konfidenční intervaly (R) ■ Korelace (R) ■ Kontingenční tabulky (R)
- Základní statistické usuzování
- Explorativní datová analýza
  - Anscombeův kvartet
- Parametrické testy hypotéz (R)
  - T-test (R) ■ F-test (R) ■ ANOVA (R) ■ MANOVA (R) ■ ANCOVA (R) ■ MANCOVA (R)
- Neparametrické testy hypotéz (R)
  - Mann-Whitneyův pořadový test (R) ■ Wilcoxonův test (R) ■ Kruskal-Wallisův test (R) ■ Friedmanův test (R)
- Chí-kvadrát testy (R)
  - Test nezávislosti (R) ■ Test dobré shody (R) ■ Reziduální analýza (R) ■ Fisherův přesný test (R)
- Metody mnohonásobného porovnání (R)
  - Tukeyův HSD test (R) ■ Nemenyi test (R)
- Resampling testy (R)

- Jacknife (R) ■ Bootstrap (R)
- Analýza síly testu (R)
- Lineární modely
  - Jednoduchá lineární regrese (R) ■ Mnohonásobná lineární regrese (R) ■ Zobecněné lineární modely (R) ■ Logistická regrese (R) ■ Analýza přežití (R) ■ Regresní diagnostika (R)
- Pokročilejší statistické metody
  - Shluková analýza (R) ■ Faktorová analýza (R) ■ Diskriminační analýza (R) ■ Časové řady (R)
- Některé metody strojového učení
  - Support Vector Machines (R) ■ Křížová validace (R) ■ Analýza principálních komponent (R) ■ Rozhodovací stromy (R) ■ Náhodné lesy (R) ■ Neuronové sítě (R)

## Grafika v R

- High-level grafické příkazy
  - Funkce plot() ■ Zobrazování mnohorozměrných dat ■ Argumenty high-level grafických příkazů
- Typy diagramů
  - Histogram (R) ■ Bodový diagram (R) ■ Sloupcový diagram (R) ■ Krabicový diagram (R) ■ Diagram s lomenou čarou (R) ■ Koláčový diagram (R) ■ Mozaikový diagram (R) ■ Korelogram (R) ■ Dendrogram (R) ■ Stem-and-leaf diagram (R) ■ Waterfall diagram (R) ■ Swimmer diagram (R)
- Low-level grafické příkazy
  - Body ■ Čáry ■ Polygony ■ Osy diagramů ■ Legenda diagramu ■ Popisky diagramu ■ Dolňkový text v diagramech ■ Argumenty low-level grafických příkazů
- Obecná grafická nastavení
  - Funkce par() ■ Barvy ■ Fonty ■ Velikosti ■ Okraje ■ Kombinace více diagramů v jednom obrázku
- Ukládání grafických výstupů na místní disk
  - Funkce png() ■ Funkce jpeg() ■ Funkce bmp() ■ Funkce tiff() ■ Funkce pdf()
- Dynamická grafika v R

## Práce s textem v R

- Základní práce s textovými řetězci
  - Funkce paste() ■ Funkce strsplit() ■ Funkce nchar() ■ Funkce substr() ■ Funkce tolower() ■ Funkce toupper()
- Regulární výrazy v R
- Sémantika regulárních výrazů
  - Kotvící značky ■ Třídy textu ■ POSIX třídy ■ Predikátové značky ■ Kvantifikátory ■ Escapeové sekvence ■ Metaznačky ■ Speciální značky ■ Skupiny a škály znaků ■ Patternové modifikátory ■ Značky řetězcových náhrad
- Některé důležité regex funkce
  - Funkce grep() ■ Funkce grepl() ■ Funkce sub() ■ Funkce gsub() ■ Funkce regexpr() ■ Funkce gregexpr() ■ Funkce regexec()
- Jednoduché příklady použití regulárních výrazů
- Pokročilejší příklady použití regulárních výrazů

## Některé užitečné balíčky

- Přehled užitečných balíčků
  - caret ■ datasets ■ devtools ■ dplyr ■ e1071 ■ foreign ■ ggplot2 ■ httr ■ igraph ■ jsonlite ■ knitr ■ lattice ■ lme4 ■ MASS ■ microbenchmark ■ OpenStreetMap ■ openxlsx ■ parallel ■ Rcpp ■ RCurl ■ rmarkdown ■ RMySQL ■ shiny ■ stringr ■ survival ■ swirl ■ tm ■ tree ■ xlsx ■ zoo

## Ukázkové příklady - vyzkoušejte si sami

### Jednodušší příklady v R

- První příklad
- Druhý příklad

### Sessions se složitější problematikou v R

- První příkladová session
- Druhá příkladová session