

Programovací jazyk R a práce s ním

Vít Gabrhel

vit.gabrhel@mail.muni.cz



**FSS MU,
19. 9. 2016**

Harmonogram

0. Organizace kurzu

1. Konceptuální představení jazyka R

2. Instalace R a Rstudio, orientace v prostředí Rstudio, úvod do psaní kódu v R

3. Základní operace – aritmetika, přiřazování proměnných, zjišťování třídy proměnných

Organizace kurzu

Syllabus:

PSY232

/

PSY532

Představení jazyka R

Co je R?

1) Programovací jazyk umožňující tvorbu "*statistického software*" (tj. analýz)

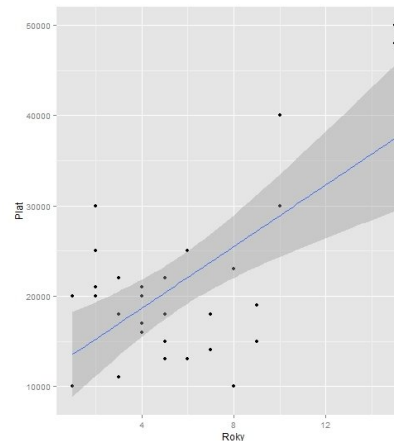
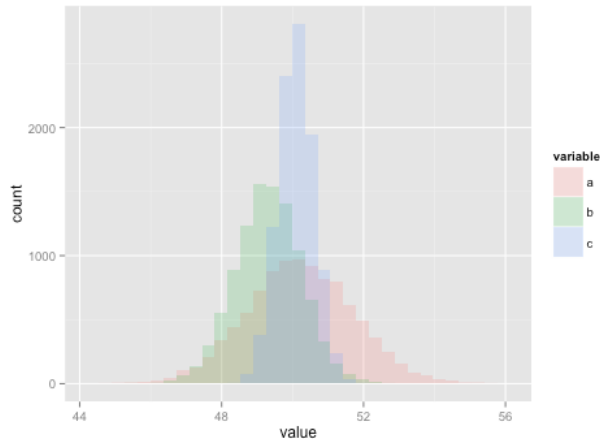
Vícenásobná lineární regrese

$$Y' = a + b_n X_n$$

$$Y' = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_n X_n + e$$

```
# Multiple Linear Regression Example
fit <- lm(y ~ x1 + x2 + x3, data=mydata)
summary(fit) # show results
```

2) Prostředí pro provádění **analýz** a **vizualizací**



Představení jazyka R

"Historie" (R (programming language), n.d.)

R je kombinací

Programovacího jazyka S určeného pro analýzu dat (Bell Labs, J. Chambers, 1976)

Lexikálního scopingu umožňujícího přiřazování dat k entitám označeným jménem

```
> košík_jablek = 1 + 2 + 4 + 8
> košík_jablek
[1] 15
```

R vyvinuli *R. Ihaka* a *R. Gentleman* z University of Auckland - první verze byla spuštěna v roce **1994**

- Název je slovní hříčkou - jednak parafrází na "S" a jednak odkazem na iniciály autorů

Představení jazyka R

Myšlenkové zázemí a současný kontext

Principy R

Open-Source a s ním související "svobody":

Svoboda používat R pro **jakékoliv účely** včetně svobody studovat to, jak R funguje a **modifikovat** ho svým vlastním potřebám.

Svoboda **šířit** R a v něm vytvořené obsahy

Svoboda **zlepšovat** R včetně **sdílení** těchto zlepšení **komunitě**.

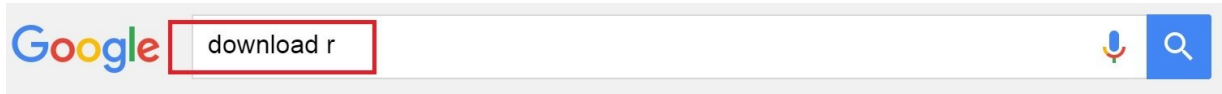
- Open source
- Transparentnost
- Komunita

Bez záruk -> větší míra odpovědnosti na uživatele

Aktuální kontext:

- Principy R a *Replikační krize*

Instalace R



All Images Videos News Books More ▾ Search tools

About 1,080,000,000 results (0.55 seconds)

[Download R-3.3.1 for Windows. The R-project for statistical computing.](https://cran.r-project.org/bin/windows/base/)

<https://cran.r-project.org/bin/windows/base/> ▾

If you want to double-check that the package you have **downloaded** exactly matches the package distributed by R, you can compare the md5sum of the .exe to ...

[R for Windows FAQ](#) · [R News](#) · [Previous Releases of R](#) · [R-3.2.1 Patched build](#)

You've visited this page 3 times. Last visit: 9/14/16

R-3.3.1 for Windows (32/64 bit)

[Download R 3.3.1 for Windows](#) (70 megabytes, 32/64 bit)

[Installation and other instructions](#)

[New features in this version](#)

If you want to double-check that the package you have downloaded exactly matches the package distributed by R, you can compare the [md5sum](#) of the .exe to the [true fingerprint](#).

Frequently asked questions

- [Does R run under my version of Windows?](#)
- [How do I update packages in my previous version of R?](#)
- [Should I run 32-bit or 64-bit R?](#)

Please see the [R FAQ](#) for general information about R and the [R Windows FAQ](#) for Windows-specific information.

Other builds

- Patches to this release are incorporated in the [r-patched snapshot build](#).
- A build of the development version (which will eventually become the next major release of R) is available in the [r-devel snapshot build](#).
- [Previous releases](#)

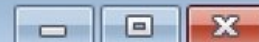
Note to webmasters: A stable link which will redirect to the current Windows binary release is <http://<CRAN MIRROR>/bin/windows/base/release.htm>.

R

File Edit View Misc Packages Windows Help

**R****R**

R Console



```
R version 3.3.1 (2016-06-21) -- "Bug in Your Hair"  
Copyright (C) 2016 The R Foundation for Statistical Computing  
Platform: i386-w64-mingw32/i386 (32-bit)
```

```
R is free software and comes with ABSOLUTELY NO WARRANTY.  
You are welcome to redistribute it under certain conditions.  
Type 'license()' or 'licence()' for distribution details.
```

```
  Natural language support but running in an English locale
```

```
R is a collaborative project with many contributors.  
Type 'contributors()' for more information and  
'citation()' on how to cite R or R packages in publications.
```

```
Type 'demo()' for some demos, 'help()' for on-line help, or  
'help.start()' for an HTML browser interface to help.  
Type 'q()' to quit R.
```

```
> "Velmi útulné"  
[1] "Velmi útulné"  
> |
```


Instalace Rstudio



download rstudio



All

Videos

Images

News

Shopping

More ▾

Search tools

About 2,280,000 results (0.44 seconds)

Download RStudio – RStudio

<https://www.rstudio.com/products/rstudio/download3/> ▾

Aug 4, 2016 - **RStudio** is a set of integrated tools designed to help you be more productive with R. It includes a console, syntax-highlighting editor that ...

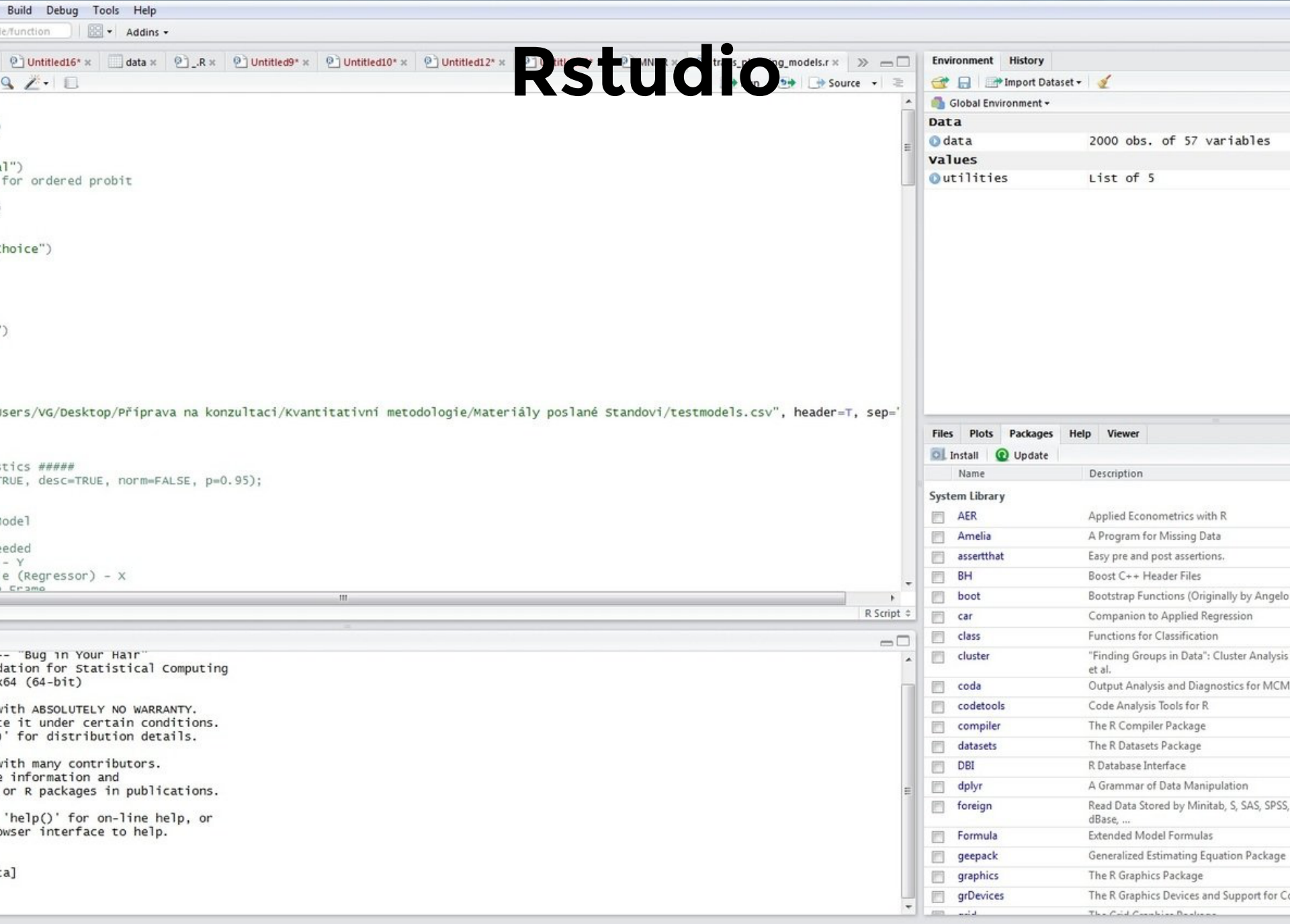
[Download RStudio Server](#) · [RStudio Release Notes](#) · [RStudio Preview](#)

Installers for Supported Platforms

Installers

Installers	Size	Date	MD5
RStudio 0.99.903 - Windows Vista/7/8/10	77.1 MB	2016-07-18	716f28f2143c5e21f4acea5752e284f8
RStudio 0.99.903 - Mac OS X 10.6+ (64-bit)	60 MB	2016-07-18	d14a1585b5a5ac0839507b9c04d460d6
RStudio 0.99.903 - Ubuntu 12.04+/Debian 8+ (32-bit)	81.6 MB	2016-07-18	761eae80b0ba4d4cd9051a802a2c44e2
RStudio 0.99.903 - Ubuntu 12.04+/Debian 8+ (64-bit)	88.3 MB	2016-07-18	98ea59d3db00e0083d3e4053514f764d
RStudio 0.99.903 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (32-bit)	81 MB	2016-07-18	ce2ea1023d99175cb909def0fe66eba7
RStudio 0.99.903 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (64-bit)	81.9 MB	2016-07-18	152f247255e86904cf3354afbc7b3b99

Rstudio



```
## SCRIPTS  
set.seed(123)  
x <- c(0,100,200,0)  
y <- c(2,3)  
z <- 105  
sales_by_month > 0]
```

Script

```
,"cat", "dog", "turtle")
```

```
2,3,4)  
[pets == "dog"]
```

```
Stats/L1 - Data management/  
mingw32/x64 (64-bit)
```

```
and comes with ABSOLUTELY NO WARRANTY.  
distribute it under certain conditions.  
'licence()' for distribution details.
```

```
project with many contributors.  
for more information and  
to cite R or R packages in publications.
```

```
See demos, 'help()' for on-line help, or  
HTML browser interface to help.
```

```
teaching/Stats/L1 - Data management")
```

Environment History

Import Dataset Clear

Global Environment

Environment is empty

Workspace

Files Plots Packages Help Viewer

New Folder Delete Rename More

Home Masaryk Teaching Stats L1 - Data management

	Name	Size
	..	
<input type="checkbox"/>	L1 R Sript draft 1.R	1.2 KB
<input type="checkbox"/>	L1 slides draft 1.pptx	1.3 MB
<input type="checkbox"/>	L1 slides draft 2.pptx	1.3 MB

Files, Help, Packages

Alternativy pro Rstudio

Graphical user interfaces [\[edit \]](#)

- Architect – cross-platform open source IDE for data science based on Eclipse and StatET
- DataJoy^[43] – Online R Editor focused on beginners to data science and collaboration.
- Deducer^[44] – GUI for menu-driven data analysis (similar to SPSS/JMP/Minitab).
- Java GUI for R – cross-platform stand-alone R terminal and editor based on Java (also known as JGR).
- Number Analytics^[4] - GUI for R based business analytics (similar to SPSS) working on the cloud.
- Rattle GUI – cross-platform GUI based on RGtk2 and specifically designed for data mining.
- R Commander – cross-platform menu-driven GUI based on tcltk (several plug-ins to Rcmdr are also available).
- Revolution R Productivity Environment (RPE) – Revolution Analytics-provided Visual Studio-based IDE, and has plans for web based point and click interface.
- RGUI – comes with the pre-compiled version of R for Microsoft Windows.
- RKWard – extensible GUI and IDE for R.
- RStudio – cross-platform open source IDE (which can also be run on a remote Linux server).

A special issue of the *Journal of Statistical Software* discusses GUIs for R.^[45]

Journal of Statistical Software (2012) - **Special Issue on Graphical User Interfaces for R:**
<https://www.jstatsoft.org/issue/view/v049>

Úvod do psaní kódu v R

Čísla a Texty

```
Ovoce = 3 + 4
```

```
Ovoce <- c("Jablko", "ČR")
```

```
names(some_vector) <- c("Ovoce", "Země původu")
```

U textové proměnné je nutné používat uvozovky ("Proč nepoužíváš Python?")

#

Mřížka či hastag **odděluje** kód od komentáře

```
# Luděk má 3 jablka, Eleonora 4 hrušky. Dohromady mají 7 kusů ovoce.
```

```
Ovoce = 3 + 4
```

Pozor na malá a velká písmena!

"=" a "->" jsou ekvivalentní formy zápisu

Úvod do psaní kódu v R

Psaní symbolů (např. "<") skrze pravý alt:

Pravý alt + další písmeno na klávesnici, např.:

- alt + . = >
- alt + f = [
- alt + v = @

Proč nefunguje příkaz $7 = 3 + 4$?

invalid (do_set) left-hand side to assignment

Parametr (například vektor nebo datová matice) nemůže v jazyce R začínat číslem. Řešení?

"7" = 3 + 4

Úvod do psaní kódu v R

Jak si "uklidit" pracovní prostor?

Příkaz `remove()`

- `remove(my_numeric)`

Jak si "říct o pomoc"?

Symbol "?" před funkcí pro vyhledávání v *dokumentaci R*

- `?remove`

Symbol "??" před funkcí pro vyhledávání příkazu *napříč knihovnamí*

- `??remove`

R jako kalkulačka

Sčítání: +

420 + 0

[1] 420

Odčítání: -

333 - 12

[1] 321

Násobení: *

2 * 3

[1] 6

Dělení: /

1 / 2

[1] 0.5

Umocňování: ^

16 ^ 2

[1] 256

Druhá odmocnina: sqrt(x)

sqrt(256)

[1] 16

Objekty

Reprezentují hodnoty (např. "0" nebo "Hruška") nebo jiné objekty (např. proměnná jako objekt v rámci datové matice, dalšího objektu).

- Skrze označení (name) je možné **vyvolat** hodnotu nebo **sadu hodnot** v tomto objektu

```
Jablka <- 3
```

```
Jablka
```

```
[1] 3
```

```
Ovoce = Jablka + Hrušky
```

```
Ovoce
```

```
[1] 7
```

```
Hrušky <- 4
```

```
Hrušky
```

```
[1] 4
```

Numerické versus **stringové** objekty

```
Hrušky = Jedna
```

```
Error: object 'Jedna' not found
```

```
Hrušky = "Jedna"
```

```
Ovoce = Jablka + Hrušky
```

```
Error in Jablka + Hrušky : non-numeric  
argument to binary operator
```

Objekty

Předchozí příklady se týkaly vyvolání hodnoty. *Co vyvolání sady hodnot?*

- Přestavte si, že plánujete uspořádat večírek a k vytvoření seznamu pozvaných se rozhodnete využít R.
- Příkaz "c()" = sloučí "argumenty" (např. 1, 0.5 nebo "Muž") do jednoho **vektoru**, tedy **sekvenci dat**

Nejdříve jejich jména:

- Pozvaní = c("Elisa", "Kai", "Edu", "Daphne")
- Pozvaní

Co vezmou s sebou?

- Dary = c("Aktivity", "Míč", "Knihu", "Lahev vína")
- Dary

A kolik toho bude?

- Množství = c(1, 3, 2, 10)
- Množství

Třídy objektů

- Decimals values like **4.5** are called **numerics**.
- Natural numbers like **4** are called **integers**.
 - Integers are also numerics.
- Boolean values (**TRUE** or **FALSE**) are called **logical**.
- **Text** (or **string**) values are called **characters**.
 - Uvozovky ("") indikují, že nějaký text je "character"

Vytvořte objekt my_numeric s hodnotou 42.5

```
my_numeric <- 42.5
```

Vytvořte objekt my_logical s hodnotou "universe"

```
my_character <- "some text"
```

Vytvořte objekt my_logical s hodnotou FALSE

```
my_logical <- FALSE
```

Typy objektů

Logical

Boolean values (**TRUE** or **FALSE**) are called **logical**.

as.logical - classifies 0 as FALSE and anything other than 0 as TRUE

- Množství = c(1, 3, 2, 10)
- as.logical(Množství)
 - *[1] TRUE TRUE TRUE TRUE*

is.logical - Create or test for objects of type "logical"

- Množství = c(1, 3, 2, 10)
- is.logical(Množství)
 - *[1] FALSE*

Třídy objektů

Jak poznat třídu dat?

Funkce **class()**:

- `class(my_numeric)`
- `class(my_character)`
- `class(my_logical)`

```
class(my_numeric)
```

```
[1] "numeric"
```

```
class(my_character)
```

```
[1] "character"
```

```
class(my_logical)
```

```
[1] "logical"
```

Factor

```
treatment <- c("A", "A", "B", "B", "Placebo", "Placebo")  
treatment.f <- as.factor(treatment)
```

```
class(treatment)
```

```
[1] "character"
```

```
class(treatment.f)
```

```
[1] "factor"
```

```
is.factor(treatment)
```

```
[1] FALSE
```

```
is.factor(treatment.f)
```

```
[1] TRUE
```

Zdroje

R (programming language). (n.d.). In Wikipedia. Staženo dne 18. 9. 2016 z
[https://en.wikipedia.org/wiki/R_\(programming_language\)](https://en.wikipedia.org/wiki/R_(programming_language))

Další čtení:

<http://www.infoworld.com/article/2940864/application-development/r-programming-language-statistical-data-analysis.html>

Cvičení

PSY232

/

PSY532

Reportujte vždy patřičnou část kódu.

Příklad zadání:

Vytvořte proměnnou `Přátelé_Jména`, která bude obsahovat hodnotu "Luděk", "Eleonora" a "Vladislav".

Příklad řešení:

```
Přátelé_Jména = c("Luděk", "Eleonora", "Vladislav")
```