

Programovací jazyk R a práce s ním

Vít Gabrhel

vit.gabrhel@mail.muni.cz



FSS MU,

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

Sylabus

Představení jazyka R

Co je R?

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

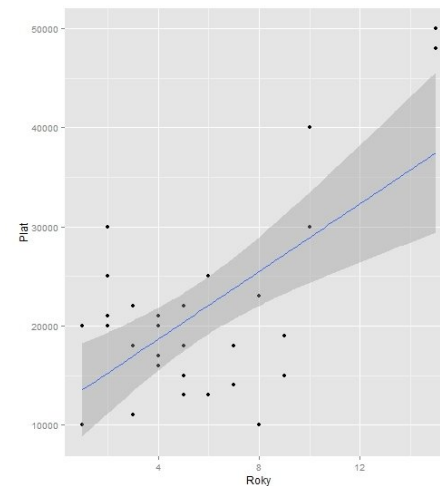
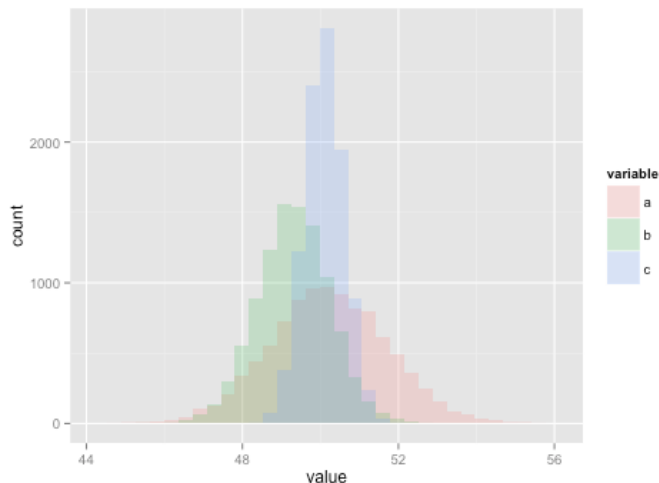
Vícenásobná lineární regrese

$$Y' = a + b_n X_n$$

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

```
# 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:

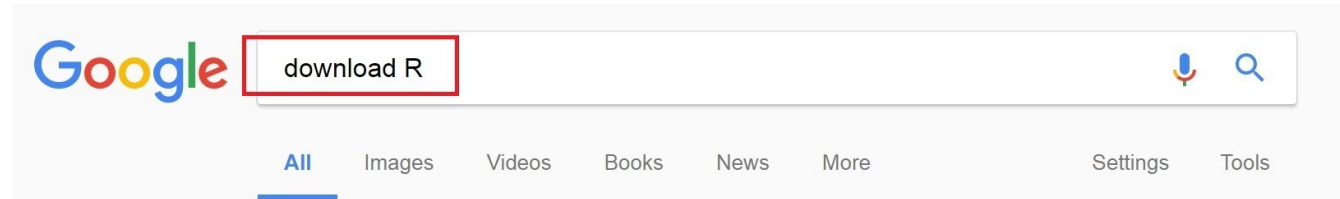
- Principy R a *Replikační krize*



You



Instalace R



About 1.780.000.000 results (0,57 seconds)

[Download R-3.4.1 for Windows. The R-project for statistical computing.](https://cran.cnr.berkeley.edu/bin/windows/base/)

<https://cran.cnr.berkeley.edu/bin/windows/base/> ▼

Download R 3.4.1 for Windows (62 megabytes, 32/64 bit) ... If you want to double-check that the package you have downloaded matches the package distributed ...

[R: The R Project for Statistical Computing](https://www.r-project.org/)

<https://www.r-project.org/> ▼

R is a free software environment for statistical computing and graphics. It compiles ... To **download R**, please choose your preferred CRAN mirror. If you have ...

[CRAN - Mirrors](#) · [Of /src/base/R-3](#) · [R: What is R?](#) · [The R Journal, Volume 9/1](#)

R-3.4.1 for Windows (32/64 bit)

[Download R 3.4.1 for Windows](#) (62 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 matches the package distributed by CRAN, you can compare the [md5sum](#) of the .exe to [graphical](#) and [command line versions](#) are available.

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.

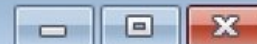
Other builds

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 | Size | Date | MD5 |
|--|---------|------------|----------------------------------|
| RStudio 1.0.153 - Windows Vista/7/8/10 | 81.9 MB | 2017-07-20 | b3b4bbc82865ab105c21cb70b17271b3 |
| RStudio 1.0.153 - Mac OS X 10.6+ (64-bit) | 71.2 MB | 2017-07-20 | 8773610566b74ec3e1a88b2fdb10c8b5 |
| RStudio 1.0.153 - Ubuntu 12.04-15.10/Debian 8 (32-bit) | 85.5 MB | 2017-07-20 | 981be44f91fc07e5f69f52330da32659 |
| RStudio 1.0.153 - Ubuntu 12.04-15.10/Debian 8 (64-bit) | 91.7 MB | 2017-07-20 | 2d0769bea2bf6041511d6901a1cf69c3 |
| RStudio 1.0.153 - Ubuntu 16.04+/Debian 9+ (64-bit) | 61.9 MB | 2017-07-20 | d584cbab01041777a15d62cbef69a976 |
| RStudio 1.0.153 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (32-bit) | 84.7 MB | 2017-07-20 | 8dfee96059b05a063c49b705eca0ceb4 |
| RStudio 1.0.153 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (64-bit) | 85.7 MB | 2017-07-20 | 16c2c8334f961c65d9bfa8fb813ad7e7 |

Rstudio

The screenshot displays the RStudio environment. At the top, the menu bar includes 'Build', 'Debug', 'Tools', and 'Help'. Below it, the toolbar contains icons for 'Run', 'Source', and other functions. The main editor window shows R code with comments and function calls. On the right, the 'Environment' pane shows the 'Global Environment' with a 'Data' section containing 'data' (2000 obs. of 57 variables) and a 'Values' section containing 'utilities' (List of 5). Below the environment pane is the 'Packages' pane, which lists installed and available packages. The 'System Library' section is expanded, showing a list of packages with their descriptions.

```
Build Debug Tools Help  
e/function Addins  
Untitled16* x data x _R x Untitled9* x Untitled10* x Untitled12* x Untitled13* x MNL.R x trans_planning_models.r x  
Run Source  
1")  
for ordered probit  
choice")  
)  
sers/VG/Desktop/Připrava na konzultaci/Kvantitativní metodologie/Materiály poslané standovi/testmodels.csv", header=T, sep=  
tics #####  
TRUE, desc=TRUE, norm=FALSE, p=0.95);  
odel  
eded  
- Y  
e (Regressor) - X  
Frame  
- "Bug in Your Hair"  
ation for Statistical Computing  
:64 (64-bit)  
with ABSOLUTELY NO WARRANTY.  
e it under certain conditions.  
' for distribution details.  
with many contributors.  
e information and  
or R packages in publications.  
'help()' for on-line help, or  
rowser interface to help.  
a]
```

| Name | Description |
|------------|---|
| AER | Applied Econometrics with R |
| Amelia | A Program for Missing Data |
| assertthat | Easy pre and post assertions. |
| BH | Boost C++ Header Files |
| boot | Bootstrap Functions (Originally by Angelo |
| car | Companion to Applied Regression |
| class | Functions for Classification |
| cluster | "Finding Groups in Data": Cluster Analysis et al. |
| coda | Output Analysis and Diagnostics for MCMC |
| codetools | Code Analysis Tools for R |
| compiler | The R Compiler Package |
| datasets | The R Datasets Package |
| DBI | R Database Interface |
| dplyr | A Grammar of Data Manipulation |
| foreign | Read Data Stored by Minitab, S, SAS, SPSS, dBase, ... |
| Formula | Extended Model Formulas |
| geepack | Generalized Estimating Equation Package |
| graphics | The R Graphics Package |
| grDevices | The R Graphics Devices and Support for Co |
| grid | The Grid Graphics Package |

```
## SCRIPTS  
set.seed(1)  
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]

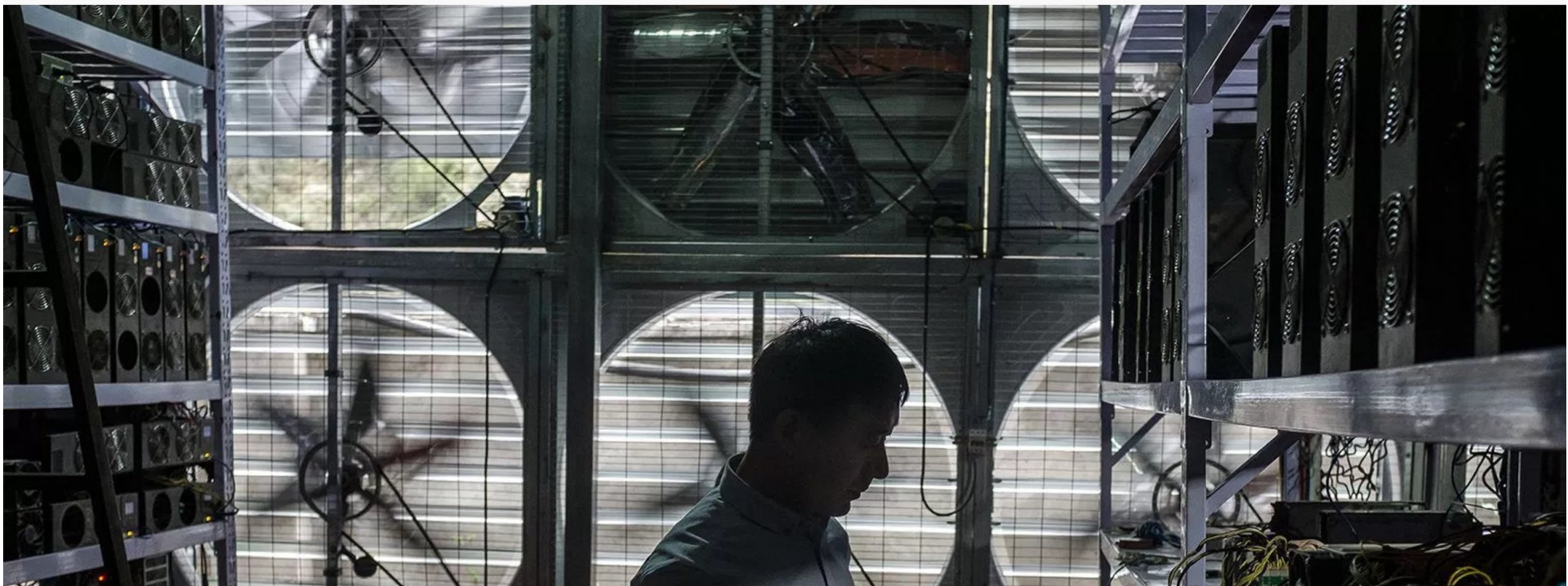
- Architect – cross-platform open source IDE for data science based on Eclipse and StatET
- DataJoy^[43] – Online R Editor focused on beginners to data science and collaboration.
- Deducer^[44] – GUI for menu-driven data analysis (similar to SPSS/JMP/Minitab).
- Java GUI for R – cross-platform stand-alone 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>

GOLDRUSH

Photos: Life inside of China's massive and remote bitcoin mines



Úvod do psaní kódu v R

Čísla a Texty

```
BitCoin = 3 + 4
```

```
CryptoCurrency <- c("Bitcoin", "Sichuan")
```

```
names(CryptoCurrency) <- c("Kryptoměna", "Provincie")
```

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

#

Mřížka či hastag **odděluje** kód od komentáře

```
# Luděk má 3 BTC, Eleonora 4 ETH. Dohromady mají 7 jednotek kryptoměn.
```

```
CryptoCurrency = 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 = @

x = 3 + 4 versus 3 + 4 = y

Error in 3 + 4 = y : target of assignment expands to non-language object

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(CryptoCurrency)`

Jak si "říct o pomoc"?

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

- `?remove`

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

- `??remove`



WHAT IF I TOLD YOU

YOU COULD USE R AS A CALCULATOR?

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

```
Bitcoin <- 3
```

```
Bitcoin
```

```
[1] 3
```

```
Kryptoměny = Bitcoin + Ethereum
```

```
Kryptoměny
```

```
[1] 7
```

Numerické versus **stringové** objekty

```
Ethereum <- 4
```

```
Ethereum
```

```
[1] 4
```

```
Bitcoin = Jedna
```

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

```
Bitcoin = "Jedna"
```

```
Kryptoměny = Bitcoin +  
Ethereum
```

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

Objekty

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

- Příkaz "c()" = sloučí "argumenty" (např. 1, 0.5 nebo "Muž") do jednoho **vektoru**, tedy **sekvenci dat**

Nejdříve vytvoříme objekt se zeměmi, kde se "těží" kryptoměny:

- Místo = c("Čína", "ČR", "GB", "Čína")
- Místo

Názvy sítí, kde se sdružují těžaři:

- Sít = c("Antpool", "Slush", "Bitclub.Network", "Bixin")
- Sít

Jaký je podíl zmíněných sítí na celku?

- Podíl = c(25, 3, 5, 8)
- Podíl

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 23.1

```
my_numeric <- 23.1
```

Vytvořte objekt my_character s hodnotou "universe"

```
my_character <- "Čína"
```

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

- Podíl = c(25, 3, 5, 8)
- as.logical(Podíl)
 - *[1] TRUE TRUE TRUE TRUE*

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

- Podíl = c(25, 3, 5, 8)
- is.logical(Podíl)
 - *[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

Zadání

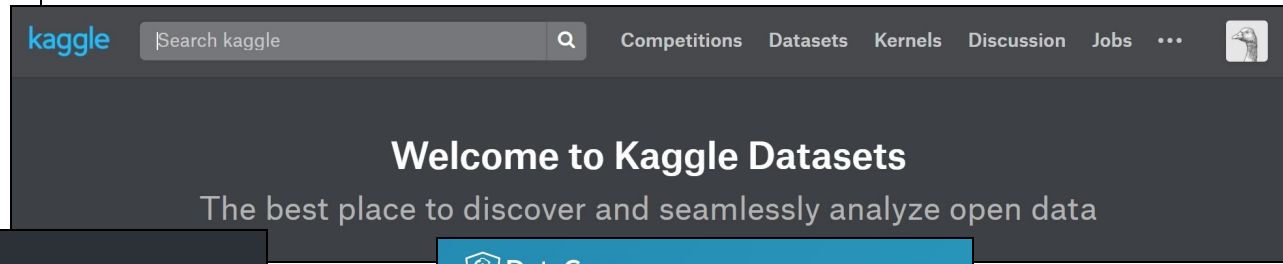
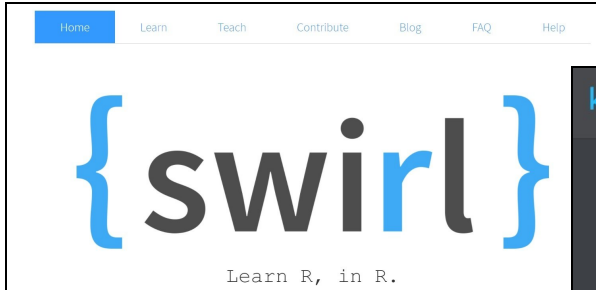
Reportujte vždy patřičnou část kódu.

Příklad řešení

Doporučené domény



Home About RSS add your blog! Learn R R jobs Contact us



Built for developers

GitHub is a development platform inspired by the way you work. From **open source** to **business**, you can host and review code, manage projects, and build software alongside millions of other developers.

The DataCamp logo is a white shield with a blue outline, containing a white icon of a person with a gear.

Learn ▾

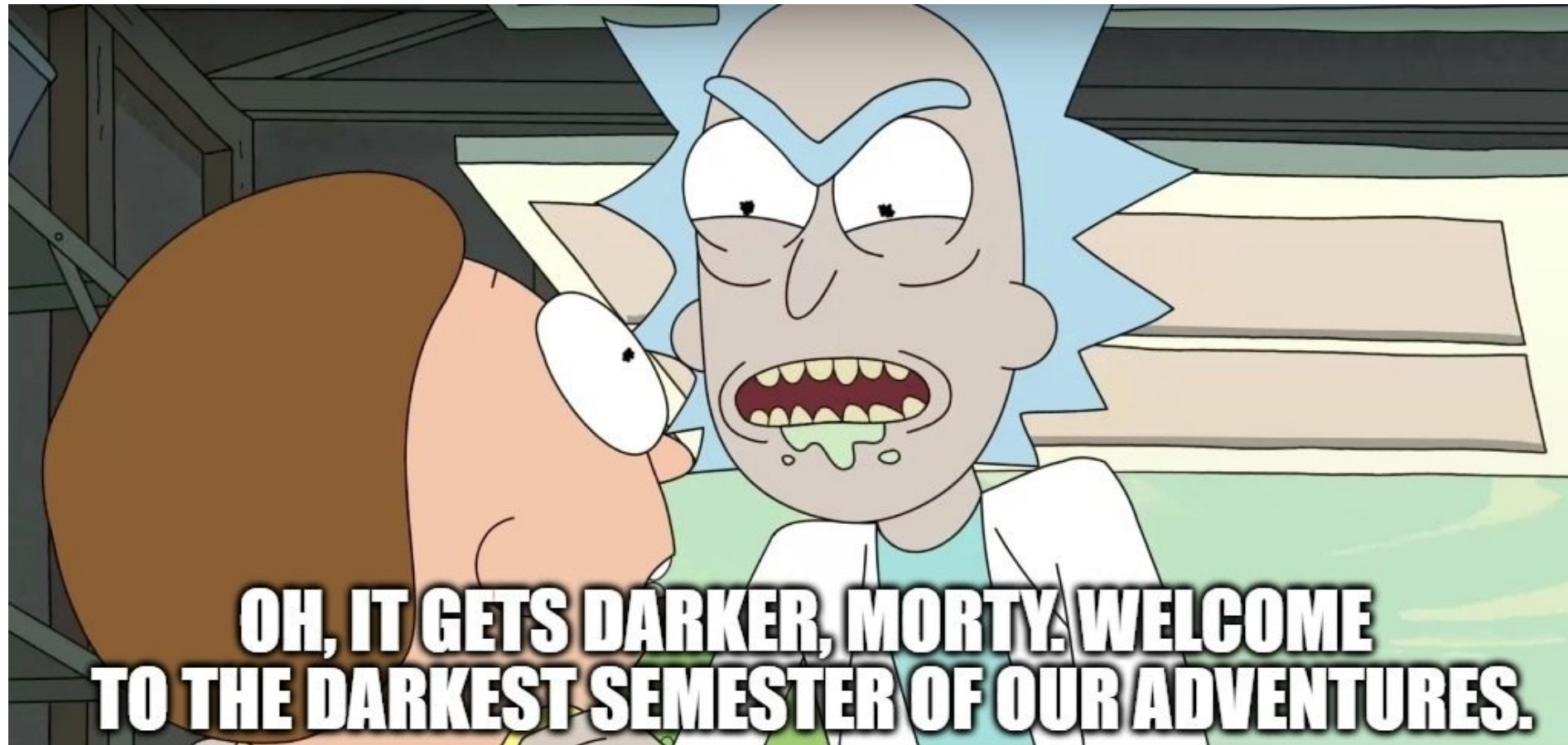
THE EASIEST WAY TO

Learn Data Science Online

Master data analysis from the comfort of your browser, at your own pace, tailored to your needs and expertise. Whether you want to learn R, Python or Data Visualization, we want to help!

Start Learning R Start Learning Python

Logos for R, Python, SQL, and Spark are displayed at the bottom of the DataCamp advertisement.



OH, IT GETS DARKER, MORTY. WELCOME TO THE DARKEST SEMESTER OF OUR ADVENTURES.