

R Markdown

Jeszenszky Péter

2022. szeptember 17.

Mi az R?

- Szabad és nyílt forrású statisztikai és grafikai nyelv és környezet.
- Programozási nyelv és interaktív környezet is egyben.
- Webhely: <https://www.r-project.org/>

Programozási nyelv

- A funkcionális, procedurális és objektum-orientált programozási paradigmákat támogatja.
- Interpretált nyelv.
 - A felhasználók által látható függvények többsége R-ben készült.
 - Számításigényes feladatok elvégzéséhez meghívásra kerülhetnek C, C++ és Fortran nyelven írt eljárások.

Népszerűség (1)

- *TIOBE Index* <https://www.tiobe.com/tiobe-index/>
 - Jelenleg a 18. legnépszerűbb programozási nyelv.
- *PYPL PopularitY of Programming Language*
<https://pypl.github.io/PYPL.html>
 - Jelenleg a 7. legnépszerűbb programozási nyelv.

Népszerűség (2)

- *2021 Kaggle Machine Learning & Data Science Survey*
<https://www.kaggle.com/c/kaggle-survey-2021>
- *The Evolution of Tools in Data Science Over the Past Five Years*
<https://www.kaggle.com/haakakak/tools-of-the-trade-a-short-history/>

R vs Python

Néhány hivatkozás a témában:

- Vik Paruchuri. *R vs Python for Data Analysis – An Objective Comparison*. October 21, 2020.
<https://www.dataquest.io/blog/python-vs-r/>
- Karlijn Willems. *Choosing R or Python for Data Analysis? An Infographic*. January 9, 2020.
<https://www.datacamp.com/community/tutorials/r-or-python-for-data-analysis>
- Ethan Schacht. *Python vs R: The Data Science Rivalry*. 2019. <https://www.kaggle.com/etsc9287/python-vs-r-the-data-science-rivalry>

Felhasználók (1)

- A felhasználók táborra nagyon vegyes.
 - A R-t sok statisztikus, biológus, közgazdász, orvos és pszichológus használja.

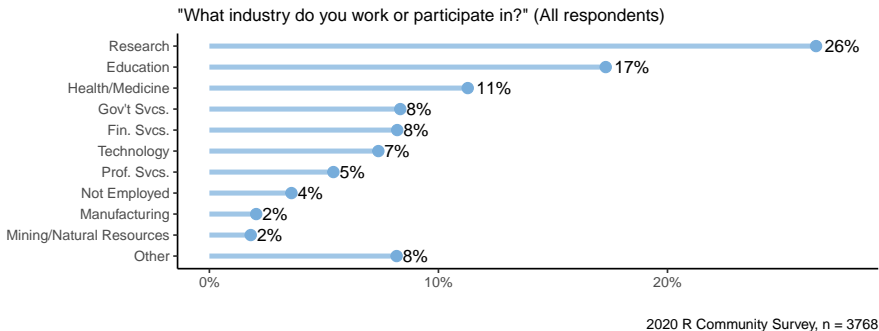
Felhasználók (2)

RStudio R Community Survey:

- Az RStudio által végzett felmérés (évek: 2018, 2019, 2020).
- Tároló: <https://github.com/rstudio/r-community-survey>

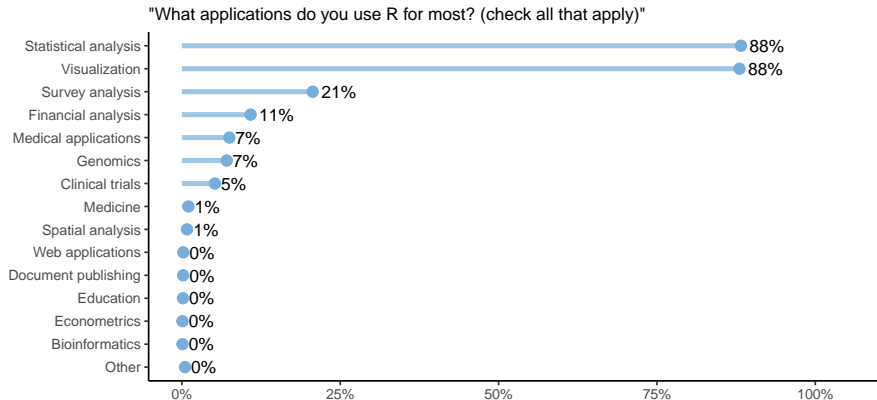
Felhasználók (3)

A felmérésben résztvevők iparági megoszlása (*RStudio R Community Survey*):



Felhasználók (4)

Az R alkalmazási területei a felmérésben résztvevők körében (*RStudio R Community Survey*):



2020 R Community Survey, n = 4028

Rendelkezésre állás (1)

- Az R a CRAN szervereken áll rendelkezésre.
 - Ugyanazt a tartalmat kínáló, földrajzilag a világ különböző részein található szerverek, melyek mindegyikén elérhető többek között:
 - A legutóbbi stabil R verzió több platformra, bináris (telepíthető) formában és forráskódban.
 - Dokumentáció
 - R csomagok
- A fő szerver a Bécsi Közgazdaságtudományi Egyetemen (*Wirtschaftsuniversität Wien*) üzemel.
- Lásd: *The Comprehensive R Archive Network*
<https://cran.r-project.org/>

Rendelkezésre állás (2)

- Több platformon is rendelkezésre áll.
 - Bináris formában tölthető le Linux, macOS és Windows rendszerekhez.
 - Más platformokra a forrásokból kell a szoftvert lefordítani és összeállítani.

Rendelkezésre állás (3)

- Az R-t a GNU GPL licenc alatt terjesztik.
 - Lásd: <https://www.r-project.org/Licenses/>
- A legtöbb csomag licence szintén a GNU GPL.
 - Esetenként előfordulhat, hogy az üzleti célú felhasználás korlátozott.
 - A nem üzleti célú felhasználás soha nincs korlátozva, azonban a GPL-nek megfelelően kell, hogy történjen.

Telepítési lehetőségek

- Bináris telepítőcsomagok ([CRAN](#))
- [Anaconda](#)/[Conda](#)
- Docker
 - [The Rocker Project](#): Docker Containers for the R Environment

Telepítés a conda segítségével

- Csomagok telepítésének engedélyezése a [conda-forge](#) csatornáról:

```
conda config --add channels conda-forge  
conda config --set channel_priority strict
```

- A legfrissebb, 4.2.1 számú verzió telepítése és használatba vétele:

```
conda create -n R-4.2.1 r-base=4.2.1  
conda activate R-4.2.1
```

R

Mi az R Markdown?

A Markdown egy kiterjesztése, mely lehetővé teszi végrehajtható programkódok beágyazását a dokumentumokba.

- Támogatott programozási nyelvek: C++, Groovy, JavaScript, R, SQL, Python, ...
- Állománynév kiterjesztés: `.Rmd`
- Webhely: <https://rmarkdown.rstudio.com/>
- Tároló: <https://github.com/rstudio/rmarkdown>
- Dokumentáció:
 - Yihui Xie, J. J. Allaire, Garrett Golemund. *R Markdown: The Definitive Guide*. 2021. <https://bookdown.org/yihui/rmarkdown/>
<https://github.com/rstudio/rmarkdown-book>

Kapcsolódó fogalmak

- Kapcsolódó programozási paradigma: [literate programming](#)
- További kapcsolódó fogalmak: reprodukálhatóság, reprodukálható kutatás (*reproducible research*)
<https://en.wikipedia.org/wiki/Reproducibility>

Alapul szolgáló szoftverek

- **rmarkdown**: R Markdown dokumentumok különféle formátumokba való alakítására szolgáló R csomag.
 - Támogatott kimeneti formátumok: HTML, PDF, PowerPoint, MS Word, ...
- **knitr**: dinamikus jelentések létrehozására szolgáló R csomag, feladata a dokumentumokban elhelyezett programkódok kiértékelése és az eredmény beillesztése.
 - Webhely: <https://yihui.name/knitr/>
 - Tároló: <https://github.com/yihui/knitr>
- **Pandoc**: a kimeneti állományok előállításához kerül felhasználásra.
 - Webhely: <https://pandoc.org/>
 - Tároló: <https://github.com/jgm/pandoc>
- **TeX rendszer**: PDF kimenet előállításához kerül felhasználásra.

Dokumentum konverzió

```
> rmarkdown::render("file.Rmd")
```

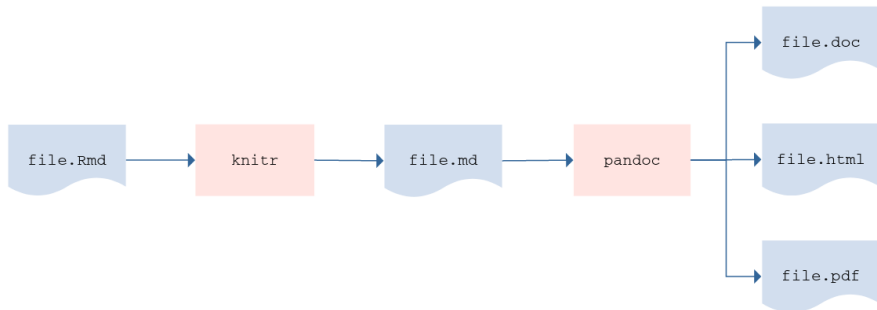


Figure 1: R Markdown Workflow

Telepítés (1)

- A Pandoc telepítése:
 - A program a `conda install pandoc` paranccsal telepíthető a `conda-forge` csatornáról.
- Az `rmarkdown` R csomag telepítése:
 - A telepítés R-ben az `install.packages("rmarkdown")` függvényhívással végezhető el.
 - A csomag a `conda-forge` csatornáról is telepíthető a `conda install r-rmarkdown` paranccsal.

Telepítés (2)

PDF kimenet előállításához $\text{T}_{\text{E}}\text{X}$ rendszer szükséges.

- Telepítés: <https://www.latex-project.org/get/#tex-distributions>
- A TeX telepítésének legegyszerűbb módja R felhasználók számára a Tiny TeX használata.

Telepítés (3)

TinyTeX:

- A TinyTeX egy pehelysúlyú, platformfüggetlen és hordozható TeX disztribúció, mely a [TeX Live](https://yihui.org/teX-live/) disztribúción alapul.
- A telepítéséhez nem szükséges adminisztrátori jogosultság.
- Webhely: <https://yihui.org/tinytex/>
- Tároló: <https://github.com/yihui/tinytex>
- Licenc: GNU GPL
- A `tinytex` R csomag szolgál a TinyTeX telepítésére, használata:

```
> install.packages("tinytex")  
> tinytex::install_tinytex()
```

Szerkesztő támogatás: RStudio Desktop

- Integrált fejlesztői környezet az R-hez.
- Webhely: <https://www.rstudio.com/products/rstudio/>
- Tároló: <https://github.com/rstudio/rstudio>
- Fejlesztő: RStudio
- Platform: Linux, macOS, Windows
- Kiadások:
 - *Open Source Edition* (licenc: GNU AGPL)
 - *RStudio Desktop Pro* (licenc: nem szabad)
- Dokumentáció: <https://docs.rstudio.com/>

Szerkesztő támogatás: Visual Studio Code

- Az *R Extension for Visual Studio Code* nyújt R és R Markdown támogatást.
 - Webhely: <https://marketplace.visualstudio.com/items?itemName=REditorSupport.r>
 - Tároló: <https://github.com/REditorSupport/vscode-R>
 - A [languageserver](#) R csomag szükséges hozzá.
 - Lehetőségek: *Knit Rmd*, *Open Preview*, *Open Preview to the Side*

Dokumentum szerkezet

Az R Markdown dokumentumok az alábbi három alapvető összetevőből állnak:

- **Metaadatok:** a dokumentum elején `---` határolók között adhatók meg YAML szintaxissal.
 - Lásd: <https://yaml.org/>
- **Szöveg:** a Pandoc Markdown eszköztára használható.
 - Lásd: *Pandoc User's Guide – Pandoc's Markdown*
<https://pandoc.org/MANUAL.html#pandocs-markdown>
- **Programkód:**
 - Hosszabb R kódrészek ````{r}` és ````` határolók között adhatók meg, ahol a kapcsos zárójelek között opciók is felsorolhatók.
 - Szövegekzi R kód ``r`` és ``` határolók között adható meg.

YAML metaadatok

- Egy jó áttekintés:
 - *YAML: an Overview* <https://cran.r-project.org/web/packages/ymlthis/vignettes/yaml-overview.html>
- Metaadatok megadása kulcs-érték párokkal `kulcs: érték` formában.
- Sztringeket szimpla vagy dupla idézőjelek között kell megadni akkor, ha speciális karaktereket (például `:` vagy `#`) tartalmaznak.
- Szövegek közötti R kód is megadható a mezőértékekben.
 - Példa: `date: '`r format(Sys.Date())`'`
- Rendelkezésre álló YAML mezők:
 - *The YAML Fieldguide* <https://cran.r-project.org/web/packages/ymlthis/vignettes/yaml-fieldguide.html>

Matematikai formulák

- Matematikai formulák LaTeX szintaxissal adhatók meg \$ vagy \$\$ határolók között.
- HTML kimenetnél a formulák megjelenítése a nyílt forrású MathJax JavaScript megjelenítő motorral történik.
 - Webhely: <https://www.mathjax.org/>
 - Tároló: <https://github.com/mathjax/MathJax>

HTML kimenet

Alapértelmezésben olyan önálló HTML dokumentumok jönnek létre, melyekbe a megjelenítéshez szükséges minden erőforrás (például szkriptek, CSS stíluslapok, képek) beágyazásra kerül.

Példa: HTML dokumentum (1)

```
---  
title: R Markdown példa  
author: Jeszenszky Péter  
date: 2021. szeptember 17.  
output: html_document  
---
```

Matematikai formula és programkód használata

Az $f(x) = \sin(\pi * x) / (\pi * x)$ függvény ábrázolása:

```
```{r}  
library(ggplot2)
f <- function(x) sin(pi * x) / (pi * x)
ggplot() + xlim(-5, 5) + geom_function(fun = f)
```
```

Példa: HTML dokumentum (2)

R Markdown példa

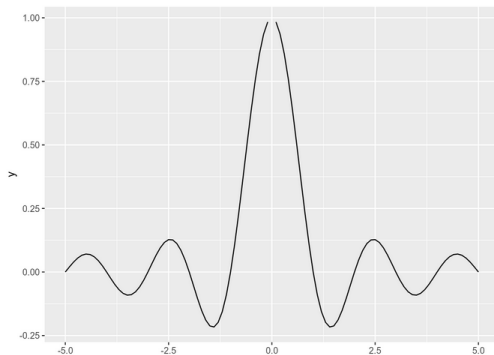
Jeszenszky Péter

2021. szeptember 17.

Matematikai formula és programkód használata

Az $f(x) = \sin(\pi * x) / (\pi * x)$ függvény ábrázolása:

```
library(ggplot2)
f <- function(x) sin(pi * x) / (pi * x)
ggplot() + xlim(-5, 5) + geom_function(fun = f)
```



Példa: PDF dokumentum (1)

```
title: R Markdown példa
author: Jeszenszky Péter
date: 2021. szeptember 17.
output: pdf_document
```

Matematikai formula és programkód használata

Az $f(x) = \sin(\pi * x) / (\pi * x)$ függvény ábrázolása:

```
```{r}
```

```
library(ggplot2)
```

```
f <- function(x) sin(pi * x) / (pi * x)
```

```
ggplot() + xlim(-5, 5) + geom_function(fun = f)
```

```
```
```

Példa: PDF dokumentum (2)

R Markdown példa

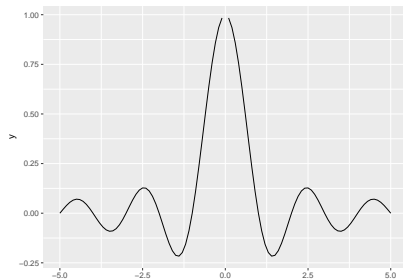
Jeszzenszky Péter

2021. szeptember 17.

Matematikai formula és programkód használata

Az $f(x) = \sin(\pi * x) / (\pi * x)$ függvény ábrázolása:

```
library(ggplot2)
f <- function(x) sin(pi * x) / (pi * x)
ggplot() + xlim(-6, 6) + geom_function(fun = f)
```



Példa: különböző formátumú kimeneti dokumentumok

- Példa:

```
---  
title: R Markdown példa  
author: Jeszenszky Péter  
date: 2021. szeptember 17.  
output:  
  html_document: default  
  pdf_document: default  
---
```

- Ilyenkor az alábbi módon állíthatók elő a kimeneti dokumentumok:

```
> rmarkdown::render("file.Rmd", output_format = "all")
```

Prezentációk (1)

- HTML prezentáció:

```
---  
output: slidy_presentation  
---
```

- További információk:

<https://bookdown.org/yihui/rmarkdown/slidy-presentation.html>

- PDF prezentáció:

```
---  
output: beamer_presentation  
---
```

- További információk:

<https://bookdown.org/yihui/rmarkdown/beamer-presentation.html>

Prezentációk (2)

- PowerPoint prezentáció:

output: powerpoint_presentation

- További információk:

<https://bookdown.org/yihui/rmarkdown/powerpoint-presentation.html>

Formátum opciók (1)

- A kimeneti állományok előállítása minden egyes formátumnál opciókkal vezérelhető, melyeket a YAML metaadat szakaszban lehet megadni.
- A rendelkezésre álló opciókat a dokumentáció tartalmazza.
 - Lásd például: `?rmarkdown::html_document`,
`?rmarkdown::pdf_document`, `?rmarkdown::beamer_presentation`

Formátum opciók (2)

Példa: formátum opciók használata HTML kimeneti dokumentumnál

output:

html_document:

code_folding: hide

df_print: paged

number_sections: true

toc: true

toc_float: true

Programozási nyelvek használata (1)

- A `knitr` csomag úgynevezett nyelvi motorok (*language engines*) révén teszi lehetővé különféle nyelvű kódrészek végrehajtását, melyek az adott nyelvű programkódot kezelő függvények.
- Az R mellett sok más nyelvhez állnak rendelkezésre nyelvi motorok.
- Bizonyos nyelvek használatához további R csomagok telepítése lehet szükséges.
 - Python: a `reticulate` csomag szükséges
 - SQL: a `DBI` csomag szükséges

Programozási nyelvek használata (2)

Rendelkezésre álló nyelvi motorok:

```
> names(knitr::knit_engines$get())
```

| | | | | | | |
|----|------|-----------|-------------|----------|----------|-----------|
| ## | [1] | "awk" | "bash" | "coffee" | "gawk" | "groovy" |
| ## | [7] | "lein" | "mysql" | "node" | "octave" | "perl" |
| ## | [13] | "Rscript" | "ruby" | "sas" | "scala" | "sed" |
| ## | [19] | "stata" | "zsh" | "asis" | "asy" | "block" |
| ## | [25] | "bslib" | "c" | "cat" | "cc" | "comment" |
| ## | [31] | "ditaa" | "dot" | "embed" | "exec" | "fortran" |
| ## | [37] | "go" | "highlight" | "js" | "julia" | "python" |
| ## | [43] | "Rcpp" | "sass" | "scss" | "sql" | "stan" |
| ## | [49] | "tikz" | "verbatim" | | | |

Programozási nyelvek használata (3)

- Bash:

```
```{bash}  
Bash kód
```
```

- Groovy:

```
```{groovy}  
// Groovy kód
```
```

- Python:

```
```{python}  
Python kód
```
```

- SQL:

```
```{sql}  
-- SQL kód
```
```


bookdown

R csomag, mely lehetővé teszi könyvek, nagyobb lélegzetű cikkek és jelentések elkészítését R Markdown formátumban.

- Webhely: <https://bookdown.org/>
<https://pkgs.rstudio.com/bookdown/>
- Tároló: <https://github.com/rstudio/bookdown>
- Dokumentáció: <https://bookdown.org/yihui/bookdown/>

A csomaggal készült könyvek listáját lásd a csomag webhelyén.

blogdown

R csomag blogok és webszajtok R Markdown dokumentumokból történő létrehozásához.

- Webhely: <https://pkgs.rstudio.com/blogdown/>
- Tároló: <https://github.com/rstudio/blogdown>
- Dokumentáció: <https://bookdown.org/yihui/blogdown/>
- Példa:
 - <https://robjhyndman.com/>

Mintapéldák

- <https://github.com/jeszy75/rmarkdown-examples/>

További információk

- Yihui Xie, Christophe Dervieux, Emily Riederer. *R Markdown Cookbook*. 2021.
<https://bookdown.org/yihui/rmarkdown-cookbook/>
<https://github.com/yihui/rmarkdown-cookbook>
- *R Markdown Cheat Sheets* <https://raw.githubusercontent.com/rstudio/cheatsheets/master/rmarkdown-2.0.pdf>