

R fyrir byrjendur

Inngangur

Anna Helga Jónsdóttir
Bjarki Þór Elvarsson
Sigrún Helga Lund

Hvað er R?

- ▶ R er umhverfi þar sem búið er að innleiða fjöldann allan af tölfræðiaðferðum
- ▶ R er open-source, allur kóði aðgengilegur
- ▶ R byggir á forritunarmálinu S
- ▶ Fyrsta útgáfa árið 2000, nýjasta útgáfa 3.5.0 (júní 2018)

Samanburður við önnur forrit

EXCEL: Ágætt til ýmissa nota en er **ekki** tölfræðihugbúnaður

SPSS: Dýrt, lokað, ...

SAS: Dýrt, lokað

MATLAB: Dýrt, lokað

Python: Líka frábært!

R for SAS and SPSS users:

www.et.bs.ehu.es/~etptupaf/pub/R/RforSAS&SPSSusers.pdf

MATLAB/R Reference:

www.math.umaine.edu/~hiebler/comp/matlabR.pdf

Kostir R

- ▶ Ókeypis!
- ▶ Fjölbreyttir gagnastrúktúrar
- ▶ Kóðinn er opinn öllum
- ▶ Mjög virkt samfélag
- ▶ ...

Ókostir R

- ▶ Krefst þolinmæði til að byrja með fyrir þá sem ekki hafa forritað áður
- ▶ Sú vinna sem þið leggið í að læra á R mun borga sig margfalt!

Hvernig er best að nota R?

- ▶ Hægt er að keyra R úr skel en betra er að nota einhvers konar ritil
- ▶ Í windows útgáfunni fylgir ritill - ekki góður
- ▶ Til eru ýmis konar kerfi/ritlar sem gerð hafa verið til að vinna með R en sá sem hefur notið mestrar hylli er RStudio (<http://rstudio.com>)
- ▶ Við munum nota RStudio hér

Náum í R og RStudio

- ▶ Þarf að byrja að sækja og setja upp R: <https://www.r-project.org/>
- ▶ Sækjum svo og setjum upp RStudio: <https://www.rstudio.com>

Keyrslugluggi, skipanagluggi og vinnusvæðið

- ▶ Skrifum annað hvort aðeins R-kóða í **skipanaskrá** (.R) eða texta og R-kóða í **skýrsluskrá** (.Rmd) sem við birtum í skipanaglugga (LaTeX fólk: .Rnw skrár)
- ▶ Allir útreikningar fara fram í keyrsluglugganum
- ▶ Í RStudio notum við `Ctrl + enter` til að senda eina línu úr skipana- eða skýrsluskrá í keyrslugluggann
- ▶ Í keyrsluglugganum er hægt er að „skrolla“ milli skipana með `↑` and `↓`
- ▶ Vinnusvæðið (workspace) inniheldur allt sem notandinn hefur búið til: vigra, fylki, föll...

Keyrslugluggi, skipanagluggi og vinnusvæðið

The screenshot shows the RStudio interface with the following components:

- Source Editor:** Contains R code for loading packages, reading data, and performing operations. The code includes comments and function calls like `library(plyr)`, `plyr::rbind.fill`, `select`, `read.table`, and `mutate`.
- Environment:** Shows the loaded data frame `minke.catch` with 85 observations and 2 variables.
- Files:** A file explorer showing the project directory with files like `C-catches.csv`, `CAT-B.DAT`, `cicmnk.dat`, `finmap.jpg`, `minkemap.pdf`, `NRMP-`, `NRMP!`, `nrmp.f`, `nrmp.R`, `NRMP.so`, `report.docx`, `report.pdf`, `report.Rmd`, `SURVEY.DAT`, and `tmp.pdf`.

Three large text overlays are present on the image:

- Skipanagluggi** (Data View): Overlaid on the Source Editor, centered over the `select` function call.
- Keyrslugluggi** (Console View): Overlaid on the Console window, centered over the output text.
- Úttakslugluggi** (Output View): Overlaid on the Environment window, centered over the data frame information.

Hvernig virkar R?

Í hvert sinn sem við framkvæmum skipun í R höfum við val um tvennt:

- ▶ **Að nota R eins og reiknivél:** skrifum skipun án þess að búa til hlut - R spýtir svarinu til baka en við getum ekki notað niðurstöðuna í framhaldinu
- ▶ **Að nota R eins og forritunarmál:** geymum það sem skipunin skilar í hlut - hluturinn er vistaður í vinnusvæðinu og við getum nálgast hann aftur og unnið frekar með hann

Hvernig virkar R?

Til að vista/búa til hlut notum við ritháttinn:

```
nafn <- skipun(arg1, arg2, ...)
```

Þar sem í stað `nafn` kemur nafnið sem við viljum gefa hlutnum og í stað `skipun` kemur nafnið á þeirri skipun sem býr til hlutinn. Nafnið má líka koma á eftir skipuninni:

```
skipun(arg1, arg2, ...) -> nafn
```

Ef við hins vegar skrifum bara skipunina, án þess að vista hana undir nafni, prentar R út það sem skipunin gerir á skjáinn, án þess að vista nokkuð.

Vinnusvæðið

- ▶ Vinnusvæðið (workspace) inniheldur allt sem notandinn hefur búið til: vigra, fylki, föll...
- ▶ R býr til skrána `.RData` sem inniheldur allt dótið
- ▶ Þegar notandinn hættir að vinna er hægt að geyma vinnusvæðið og hlaða því svo inn þegar notandinn hefur vinnu
- ▶ Auðvelt að flytja/deila `.RData`

Vinnusvæðið - gagnlegar skipanir

```
getwd()           # staðsetning möppu sem  
                  # unnið er í  
setwd()           # breytir staðsetningu  
rm(hlutur)        # fjarlægir hlutinn hlutur  
save(hlutur, 'nafn.RData') # geyma hlut  
load('nafn.RData') # hlaða inn hlut  
q()               # hætta, R spyr hvort  
                  # geyma eigi vinnusvæði
```

Fyrir windows notendur

R skilur ekki slóð á eftirfarandi formi:

```
c:\mydocuments\verk
```

Þarf að vera á eftirfarandi formi:

```
c:\\mydocuments\\verk
```

eða

```
c:/mydocuments/verk
```

setwd skipunin gæti því litið svona út:

```
setwd("c:/mydocuments/verk")
```

Verkefnautanumhald í RStudio

- ▶ Þegar unnið er að fleiri en tveim úrvinnslum í einu og/eða við viljum koma að sömu úrvinnslu síðar meir getur RStudio auðveldað okkur lífið með því að skilgreina verkefni (projects)
- ▶ Verkefni eru sérstakar skrár með endingunni .Rproj. Þær halda utan um vinnumöppu notandans, tryggja að ný (og hrein) vinnulota er ræst og opnar þær skrár sem notandinn var að vinna með síðast
- ▶ Nýtt verkefni er stofnað með því að velja File og New project og svo í framhaldinu velja viðeigandi hluti í valglugga sem þá birtist
- ▶ Opna má verkefni með að velja File og Open project

Hjálp

```
help(fall)           # hjálp fyrir fallid 'fall'
?fall                # sama
help.start()         # almenn hjálp
help.search('efni')  # leitar að 'efni' í hjálp
example(fall)        # dæmi um notkun fallisins
apropos('str')       # listi yfir föll með 'str'
                     # í nafni
RSiteSearch('fall')  # leita að 'fall' í hjálp
                     # og póstlistum
```


Íslenskir stafir

- ▶ Íslenskir stafir eru oft til vandræða þegar skipanaskrár eru opnuð í annarri gerð af stýrikerfi en þær voru búnar til í
- ▶ Þá er mikilvægt að vista ekki skjalið heldur fara beint í File, Reopen with Encoding og velja þá:
 - ▶ ISO-8859-1 ef unnið er í MacOSX eða Linux tölvu en skjalið var búið til í Windows tölvu.
 - ▶ UTF-8 ef unnið er í Windows tölvu en skjalið var búið til í MacOSX eða Linux tölvu
- ▶ Ef þið hyggist vinna áfram með skjalið borgar sig svo að fara í File, Save with Encoding og velja þar þann staðal sem tölvun ykkar notar

Íslenskir stafir

```
← → |  ☐ Source on Save |   |  |  Run |  Source ▾  
1 #lesum inn gæðgn  
2 data <- read.table("puls.csv", header=TRUE, sep=';')  
3  
4 #skoðum gæðgnin  
5 str(data)  
6  
7 #reiknum helstu lágstærðir fyrir breytuna haed:  
8 mean(data$haed)  
9 median(data$haed)  
10 min(data$haed)  
11 max(data$haed)  
12 sd(data$haed)  
13  
14 #summary  
15 summary(data$haed)  
16:1 | (Top Level) ⇅ R Script ⇅
```

Íslenskir stafir

Einnig lenda MacOSX notendur stundum í því að íslenskir stafir birtast rétt í .R skráum en rangt á gröfum. Það er hægt að laga með því að gefa skipunina:

```
system("defaults write org.R-project.R force.LANG en_US.UTF-8")
```

Þessa skipun á aðeins að keyra ef þið eruð með Makka og eruð að lenda í vandræðum með íslenska stafi á gröfum!

Pakkar í R

- ▶ Einn helsti styrkleiki R eru fjöldamargir notendur sem stöðugt búa til nýja pakka
- ▶ Notendur hlaða pökkum upp á CRAN (Comprehensive R Archive Network)
- ▶ Fjöldi pakka aðgengilegur á CRAN 14. júní 2018: 12633
- ▶ Öllum pökkum fylgir .pdf skjal með leiðbeiningum

Pakkar í R - leit

- ▶ Lista yfir pakka, raðað eftir dagsetningu má finna á:
<http://mirrors.dotsrc.org/cran/> (smellið á packages)
- ▶ Task views má finna á:
<http://cran.r-project.org/web/views/>
- ▶ Crantastic! <http://crantastic.org/>
- ▶ Mailing list: <http://www.r-project.org/mail.html>
- ▶ Samansafn frábærra pakka: tidyverse <https://www.tidyverse.org/>

Hlaða inn pökkum í R

- ▶ Við notum `install.packages()` skipunina til að sækja og hlaða inn pökkum - þarf aðeins að gera einu sinni
- ▶ Í hvert skipti sem við ætlum að nota pakka þarf að hlaða honum inn með `library()` skipuninni
- ▶ Ef verið er að vinna í RStudio má smella á Packages
- ▶ Notandi þarf að hafa skrifréttindi í R-möppunni...

```
install.packages("tidyverse") # einu sinni  
library(tidyverse) # í hverri vinnulotu
```

Ýmist gott efni um R

- ▶ <http://r4ds.had.co.nz/>
- ▶ <http://www.cookbook-r.com/>
- ▶ <https://www.statmethods.net/>
- ▶ Googla!