

levads darbā ar R un RStudio

Didzis Elferts

Lekcijas tēmas

- Par programmu R
- Komandu veidošana
- Datu importēšana
- Darbs ar datiem

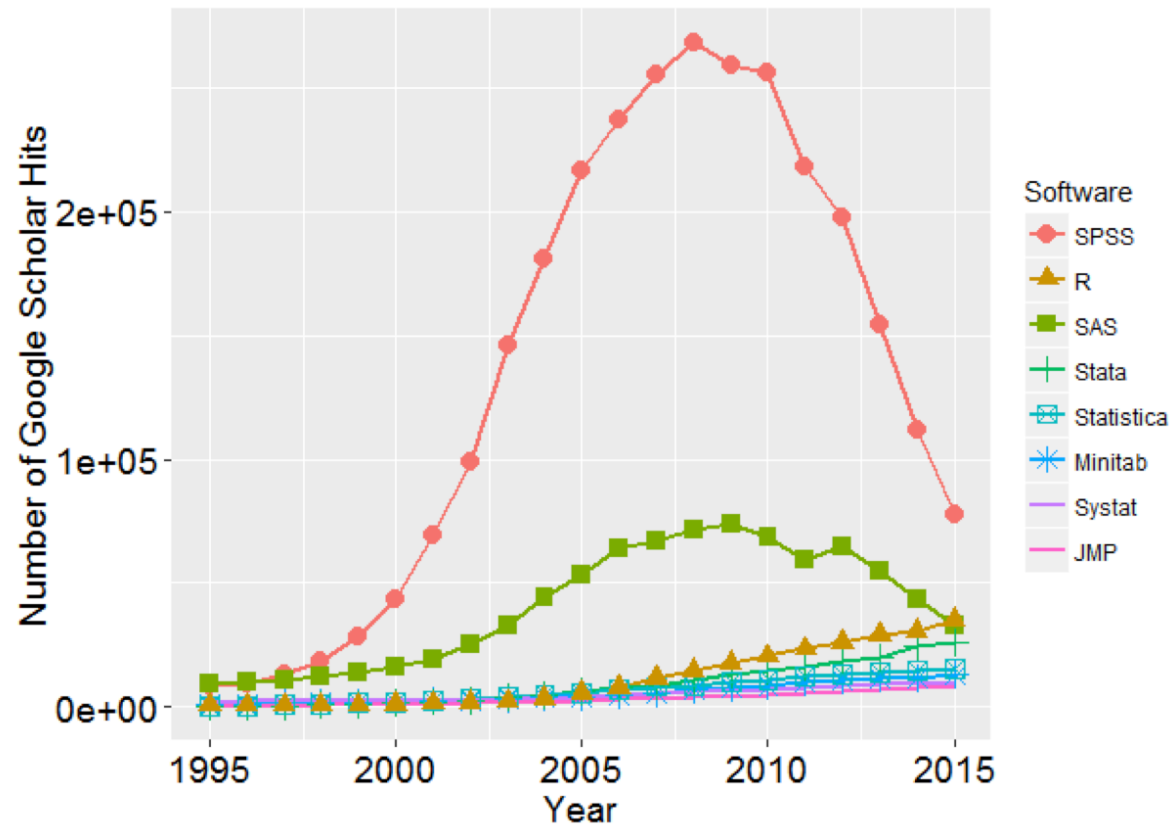
R priekšrocības

- Atvērtā koda programma, kas nepārtraukti tiek papildināta un uzlabota
- R var tikt izmantots uz dažādām platformām: Linux, Windows, MacOS
- Lielākā daļa statistisko metožu un aprēķinu ir iekļauts šajā programmā
- Lietotājam ir iespējams kontrolēt visus parametrus veicot aprēķinus
- Ļoti labas grafiskās iespējas, kas ļauj veidot augstas kvalitātes grafikus
- Iespēja veidot pašam savas funkcijas
- Var nodrošināt reproducējamību (reproducible)

R trūkumi

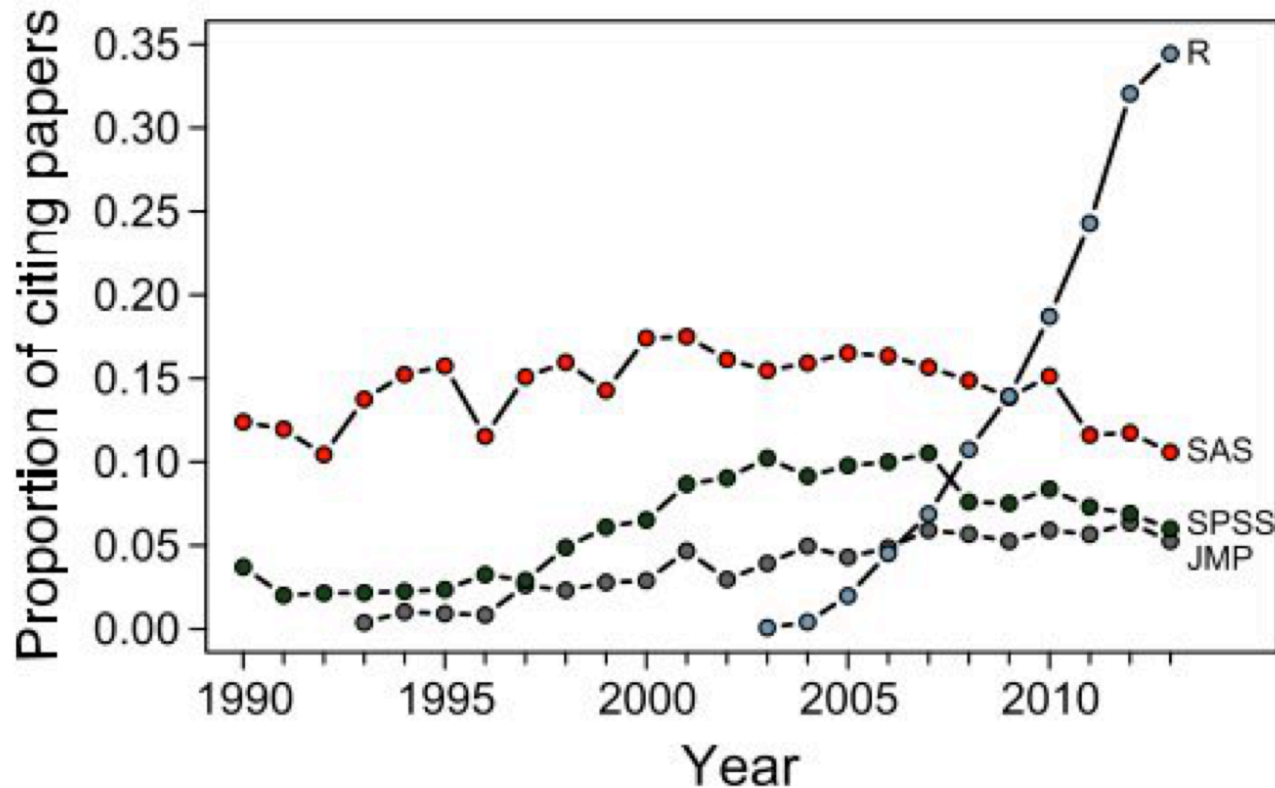
- Lēns "mācīšanās" temps
- Daļējs trūkums - komandu rindas
- Grūti izsekot līdz visām papildus paketēm
- Ne vienmēr var ātri atrast nepieciešamo

R popularitate



Avots: <http://r4stats.com/articles/popularity/>

Popularity of R in ecology



Avots: <http://onlinelibrary.wiley.com/doi/10.1002/ecs2.1394/full>

Touchon, J.C., McCoy, M.W., 2016. The mismatch between current statistical practice and doctoral training in ecology. *Ecosphere*, 7 (8).

Piemērs darbā ar R

Izveidojiet Excel failu, kurā ir divas kolonnas - Garums un Svars, ierakstiet šajās kolonnās izgudrotas 5-10 rindiņas. Saglabājiet šo failu ar nosaukumu Cilveki.xlsx direktorijā My documents (Mani dokumenti)

Piemērs darbā ar R

```
if (!require("readxl") ) install.packages("readxl")
library(readxl)
dati <- read_excel("Cilveki.xlsx")
dati
```

```
## Loading required package: readxl
```

```
## # A tibble: 6 × 2
```

```
##   Garums Svars
```

```
##   <dbl> <dbl>
```

```
## 1    175    70
```

```
## 2    180    86
```

```
## 3    165    61
```

```
## 4    163    69
```

```
## 5    172    72
```

```
## 6    169    68
```


Programmas

Programmu R var lejupielādēt no mājaslapas <https://www.r-project.org>, izraugoties CRAN serveri un savai operētājsistēmai atbilstošo versiju.

Programmas RStudio bezmaksas versiju var lejupielādēt:

- jaunākā stabilā versija <https://www.rstudio.com/products/rstudio/download/>
- jaunākā izstrādes versija <https://www.rstudio.com/products/rstudio/download/preview/>

Piemērs

```
cor.test(dati$Garums,dati$Svars)
```

```
##  
## Pearson's product-moment correlation  
##  
## data:  dati$Garums and dati$Svars  
## t = 2.8434, df = 4, p-value = 0.04671  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
##  0.01894518 0.97937919  
## sample estimates:  
##          cor  
## 0.8179306
```

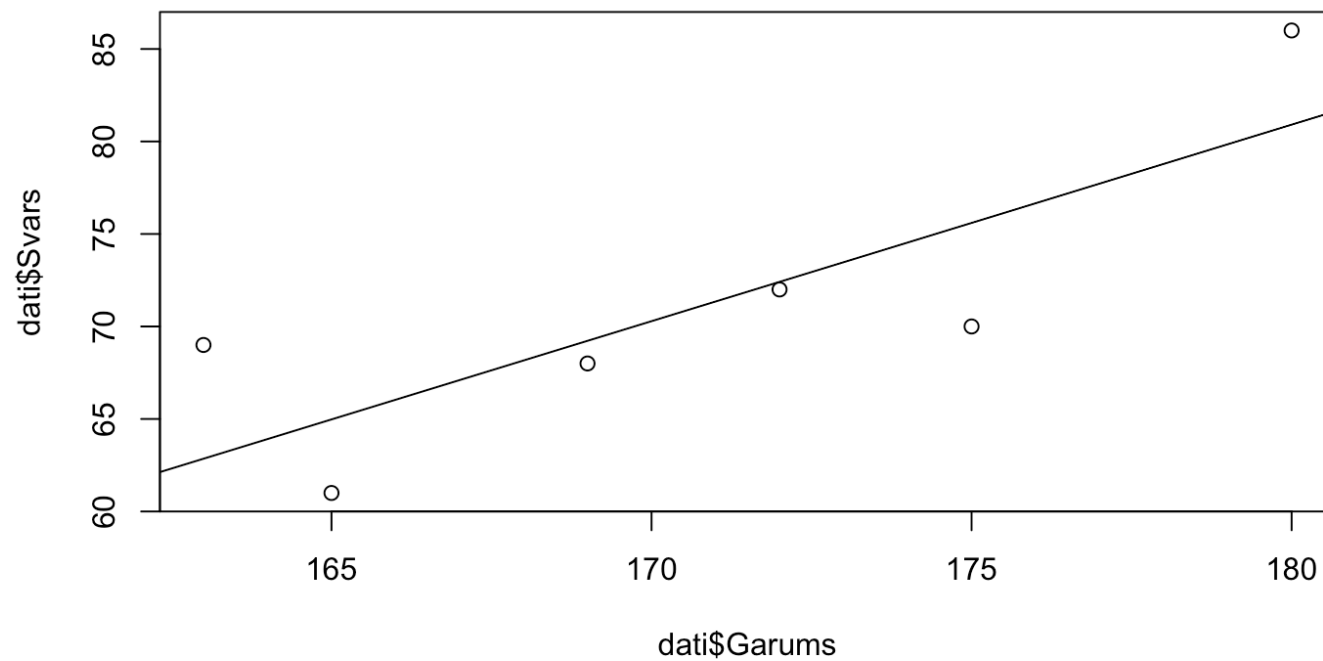
Piemērs

```
summary(lm(Svars~Garums,data=dati))
```

```
##  
## Call:  
## lm(formula = Svars ~ Garums, data = dati)  
##  
## Residuals:  
##      1      2      3      4      5      6  
## -5.6060  5.0795 -3.9768  6.1490 -0.4172 -1.2285  
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)  
## (Intercept) -110.4040    63.8348  -1.730   0.1588  
## Garums       1.0629     0.3738   2.843   0.0467 *  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 5.304 on 4 degrees of freedom  
## Multiple R-squared:  0.669, Adjusted R-squared:  0.5863  
## F-statistic: 8.085 on 1 and 4 DF, p-value: 0.04671
```

Piemērs

```
plot(dati$Svars~dati$Garums)  
abline(lm(Svars~Garums,data=dati))
```



Komandu veidošana

- Lai komandu rindai pievienotu komentāru, pirms tā ir jāraksta “#”
- Atstarpes komandās parasti tiek ignorētas, izņēmums ir rakstot “<-”
- Ja komanda ir pārāk gara, tad to var sadalīt vienkārši ar Enter taustiņu
- Iztrūkstošās vērtības programmā R apzīmē ar NA

Palīdzības iegūšana

```
help.start()
```

```
help(plot)
```

```
args(cor)
```

```
## function (x, y = NULL, use = "everything", method = c("pearson",  
##           "kendall", "spearman"))  
## NULL
```

```
example(plot)
```

Mājaslapā:

<http://www.rdocumentation.org/>

R paketes

- Bāzes R programmā ir tikai neliela daļa no iespējām, ko piedāvā R
- Papildus iespējas (funkcijas) pieejamas kā paketes (library), kuras no sākuma jāuzinstalē (ja tas jau nav izdarīts) un tad jāpievieno (tas jādara katrā sesijā)
- Pakešu instalācija notiek ar funkciju `install.packages()`, bet pievienošana darba sesijai ar funkciju `library()`.

```
install.packages("ggplot2")  
library(ggplot2)
```

R kā kalkulators

R var tikt izmantots kā vienkāršs kalkulators

```
4+7
```

```
## [1] 11
```

```
log(8,2)
```

```
## [1] 3
```

```
exp(2)
```

```
## [1] 7.389056
```


R citēšana

Ja datu analīzei izmanto R/RStudio, tad darbā obligāti jānorāda atsauce uz programmu R (arī tad, kad izmantojat RStudio, reāli strādājat ar R). Tekstā jānorāda arī R versija.

```
citation()
```

```
##
```

```
## To cite R in publications use:
```

```
##
```

```
## R Core Team (2016). R: A language and environment for
```

```
## statistical computing. R Foundation for Statistical Computing,
```

```
## Vienna, Austria. URL https://www.R-project.org/.
```

```
##
```

```
## A BibTeX entry for LaTeX users is
```

```
##
```

```
## @Manual{,
```

```
## title = {R: A Language and Environment for Statistical Computing},
```

```
## author = {{R Core Team}},
```

```
## organization = {R Foundation for Statistical Computing},
```

```
## address = {Vienna, Austria},
```

R citēšana

Ja izmantota kāda papildus pakete, tad jānorāda atsauce arī uz to.

```
citation("readxl")
```

```
##  
## To cite package 'readxl' in publications use:  
##  
## Hadley Wickham (2016). readxl: Read Excel Files. R package  
## version 0.1.1. https://CRAN.R-project.org/package=readxl  
##  
## A BibTeX entry for LaTeX users is  
##  
## @Manual{,  
##   title = {readxl: Read Excel Files},  
##   author = {Hadley Wickham},  
##   year = {2016},  
##   note = {R package version 0.1.1},  
##   url = {https://CRAN.R-project.org/package=readxl},  
## }
```

Interaktīva R macīšanās

Izmantojot paketi `swirl` ir iespējams interaktīvi mācīties strādāt programmā R (angļu valodā)

```
library(swirl)
```

Datu importēšana

- Bāzes R piedāvā iespējas importēt .txt, .csv un citus ASCII tipas datus
- Izmantojot papildus paketes iespējams importēt ļoti daudzus citus datu formātus - Excel, SPSS, SAS, json, utt

Datu importēšana

```
read.table(file="/../faila.nosaukums.txt", header=TRUE, sep="\t", dec=".")
```

Kolonnu atdalītāja `sep` vērtības var būt: `"\t"` , `" "` ; `"`

Decimāldaļu atdalītājs failā `dec` varbūt `"."` vai `","`

Pēc noklusējuma visas teksta kolonnas tiek pārvērstas par faktoriem. To var novērst pievienojot argumentu `stringsAsFactors=FALSE`.

Garo saiti uz failu var nerakstīt, ja fails atrodas darba direktorijā (Working directory) (var nomainīt ar `File/Change dir...`)

Darba direktorija

```
getwd()
```

```
## [1] "/Users/didzis/Dropbox/DidzaDati/2016/R_kursi_LU/R_kursi_LU/Ievads_R_RStudio"
```

```
setwd("/Users/didzis/Documents")
```

Datu importēšana

```
dati <- read.table(file="niedres.txt",header=T,  
                  sep="\t", dec=".")  
dati2 <- read.csv2(file="niedres.txt",header=T,  
                  sep="\t", dec=".")
```

dati

##	garums	platums
## 1	31.6	2.5
## 2	23.2	2.3
## 3	39.2	2.1
## 4	37.4	5.8
## 5	21.1	2.2
## 6	37.0	4.1
## 7	24.7	3.5
## 8	31.3	4.2
## 9	37.4	2.5
## 10	39.7	2.8
## 11	38.0	4.3
## 12	24.9	1.1

Datu importēšana - iebūvētie dati

```
data()
```


Datu importēšana - iebūvētie dati

```
data(cars)
```

```
head(cars)
```

```
##      speed dist  
## 1         4    2  
## 2         4   10  
## 3         7    4  
## 4         7   22  
## 5         8   16  
## 6         9   10
```

Datu eksportēšana

Datus, kas izveidoti R kā datu tabulas, visērtāk var eksportēt kā `.txt` vai `.csv` failus.

```
write.table(x=tas.kas.jaeksporte,  
  file="/saite/uz/failu/failanosaukums.txt",  
  sep="\t",dec=".")
```

```
write.table(x=dati, file="eksports.txt",sep="\t",dec=".")
```

Eksportējot pēc noklusējuma tiek eksportēti arī rindiņu nosaukumi. Ja tos nevajag, tad jāpievieno arguments `row.names=FALSE`

Lietas, kas jāatceras

- Programmā R visi radītie, importētie objekti atrodas operatīvajā atmiņā (ja neizmanto specializētas pieejas)
- Ja neizmanto speciālas paketes, tad R darbībā tiek izmantots tikai viens datora procesors

Datu veidi

- Skaitliski (numeric), piemēram, 1, 5, 2.6, -123.45
- Rakstu zīmes (character), piemēram, AA, zaļš, mētrājs
- Loģiski (logical) - TRUE un FALSE

Datu struktūras

- Vektors (vector)
- Matrice (matrix)
- Saraksts (list)
- Datu rāmis/tabula (data frame)
- Datu tabula (data table)
- u.c.

Datu atlasīšana – datu tabula

```
dati$garums
```

```
## [1] 31.6 23.2 39.2 37.4 21.1 37.0 24.7 31.3 37.4 39.7 38.0 24.9 30.8 26.7
## [15] 34.7 33.3 59.0 17.7 24.3 41.4 49.1 46.2 11.9 39.5 45.3 39.6 45.8 31.0
## [29] 39.9 56.4 37.3 41.1 45.3 49.3 24.9 51.7 40.6 43.5 48.9 31.6 53.6 53.5
## [43] 26.9 35.7 46.1 29.1 33.7 43.1 38.8 46.9
```

```
dati["garums"]
```

```
##      garums
## 1      31.6
## 2      23.2
## 3      39.2
## 4      37.4
## 5      21.1
## 6      37.0
## 7      24.7
## 8      31.3
## 9      37.4
```

Datu atlasīšana – datu tabula

Otrais indekss kvadrātiskajās iekavās nozīmē kolonnas numuru.

```
dati[,1]
```

```
## [1] 31.6 23.2 39.2 37.4 21.1 37.0 24.7 31.3 37.4 39.7 38.0 24.9 30.8 26.7
## [15] 34.7 33.3 59.0 17.7 24.3 41.4 49.1 46.2 11.9 39.5 45.3 39.6 45.8 31.0
## [29] 39.9 56.4 37.3 41.1 45.3 49.3 24.9 51.7 40.6 43.5 48.9 31.6 53.6 53.5
## [43] 26.9 35.7 46.1 29.1 33.7 43.1 38.8 46.9
```

```
dati[,1,drop=FALSE]
```

```
##      garums
## 1      31.6
## 2      23.2
## 3      39.2
## 4      37.4
## 5      21.1
## 6      37.0
## 7      24.7
```

Datu atlasīšana – datu tabula

Pirmais indekss kvadrātiskajās iekavās nozīmē rindas numuru.

```
dati[1,]
```

```
##   garums platums  
## 1   31.6    2.5
```

```
dati[c(2,5),]
```

```
##   garums platums  
## 2   23.2    2.3  
## 5   21.1    2.2
```

```
dati[1:3,]
```

```
##   garums platums  
## 1   31.6    2.5  
## 2   23.2    2.3
```


Darbs ar datiem

```
garums
```

```
Error: object 'garums' not found
```

Programmā R datu tabulu kolonnas automātiski netiek uzskatītas par atsevišķiem mainīgkiem

Darba vides sakārtošana

Atmiņā esošo objektu apskatīšana

```
ls()
```

```
## [1] "cars" "dati" "dati2" "img" "saite" "z"
```

Objekta dzēšana

```
rm(dati)
```