

R

Pasniedzējs

Didzis Tjarve

E-pasts

Didzis.Tjarve@lu.lv

Programmatūra datu ievadei un organizēšanai

- Teksta redaktori – Word, OpenOffice Writer, Notepad, Notepad ++
- Elektronisko tabulu redaktori – Excel, OpenOffice Calc
- Datu bāzu vadības sistēmas – Access, Oracle, MySQL, SQLite

Teksta redaktori

- Ērti veikt korekcijas, ja dati glabāti teksta failos
- Ērti aplūkot datus
- Samērā sarežģīta liela apjoma datu ievade
- Ļoti ierobežotas datu pārorganizēšanas un atlasas iespējas

Elektronisko tabulu redaktori

- Ērta datu ievade
- Plašas datu pārorganizēšanas iespējas
- Ierobežotas datu atlases iespējas
- Ierobežots ievadāmo datu apjoms

Datu bāzu vadības sistēmas

- Nav ievadāmo datu apjoma ierobežojumu
- Ļoti augstas datu atlašes iespējas
- Neaizstājamas liela datu apjoma gadījumā
- Datu pārorganizēšana un atlaše var prasīt speciālas iemaņas
- Programmētājs var veidot konkrētā lietotāja vajadzībām pielāgotu sistēmu

Datu failu formāti

- XLS, ODS
- DBF, Accdb
- CSV, TXT

Teksta faili

- atdalītājs komats vai semikols (*comma delimited*)
 - saglabājot *Excel* programmā atdalītājs ir tas, kas norādīts sistēmas iestādnēs (*Regional settings*) kā *List separator*
- atdalītājs tabulators (*tab delimited*)

Piemērs

- Programmā *Excel* atvērt failu *piemers_1.xls*
- Failu saglabāt kā
 - Text (Tab delimited)
 - CSV (Comma delimited)
- Failus atvērt programmā *Notepad ++* un noskaidrot atdalītājus

Datu analīzes programmas

- Komerčiālās
 - SPSS
 - Statgrafic
- Atvērtā koda
 - Specializētas
 - R

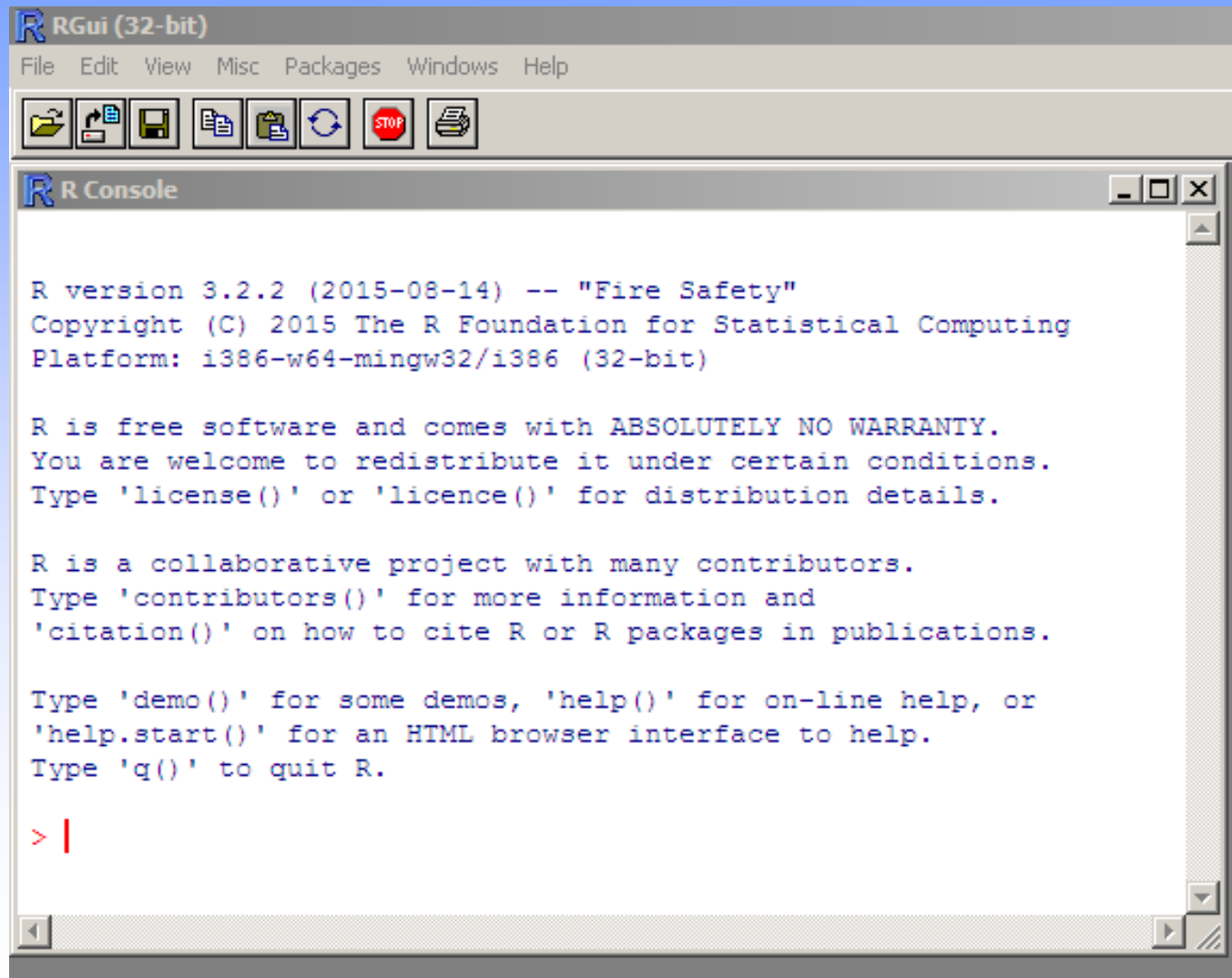
R

- <https://www.r-project.org/>
- Atvērtā koda programma, kas pastāvīgi tiek uzlabota
- Var izmantot dažādās vidēs – Windows, Linux, Mac OS

R iespējas

- Pieejams liels skaits funkciju bibliotēku, kas aptver praktiski visas statistiskās metodes
- Iespēja veidot ļoti augstas kvalitātes grafikus
- Pārvarā darbs ar komandrindu
- Pieejams arī «*R Commander*», kas nodrošina darbu ar izvēlnēm

R vide



The image shows a screenshot of the RGui (32-bit) application window. The title bar reads "RGui (32-bit)". The menu bar includes "File", "Edit", "View", "Misc", "Packages", "Windows", and "Help". Below the menu bar is a toolbar with icons for file operations: Open, Save, Print, Copy, Paste, Refresh, Stop, and Print. The main window contains an "R Console" pane with the following text:

```
R version 3.2.2 (2015-08-14) -- "Fire Safety"
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: i386-w64-mingw32/i386 (32-bit)

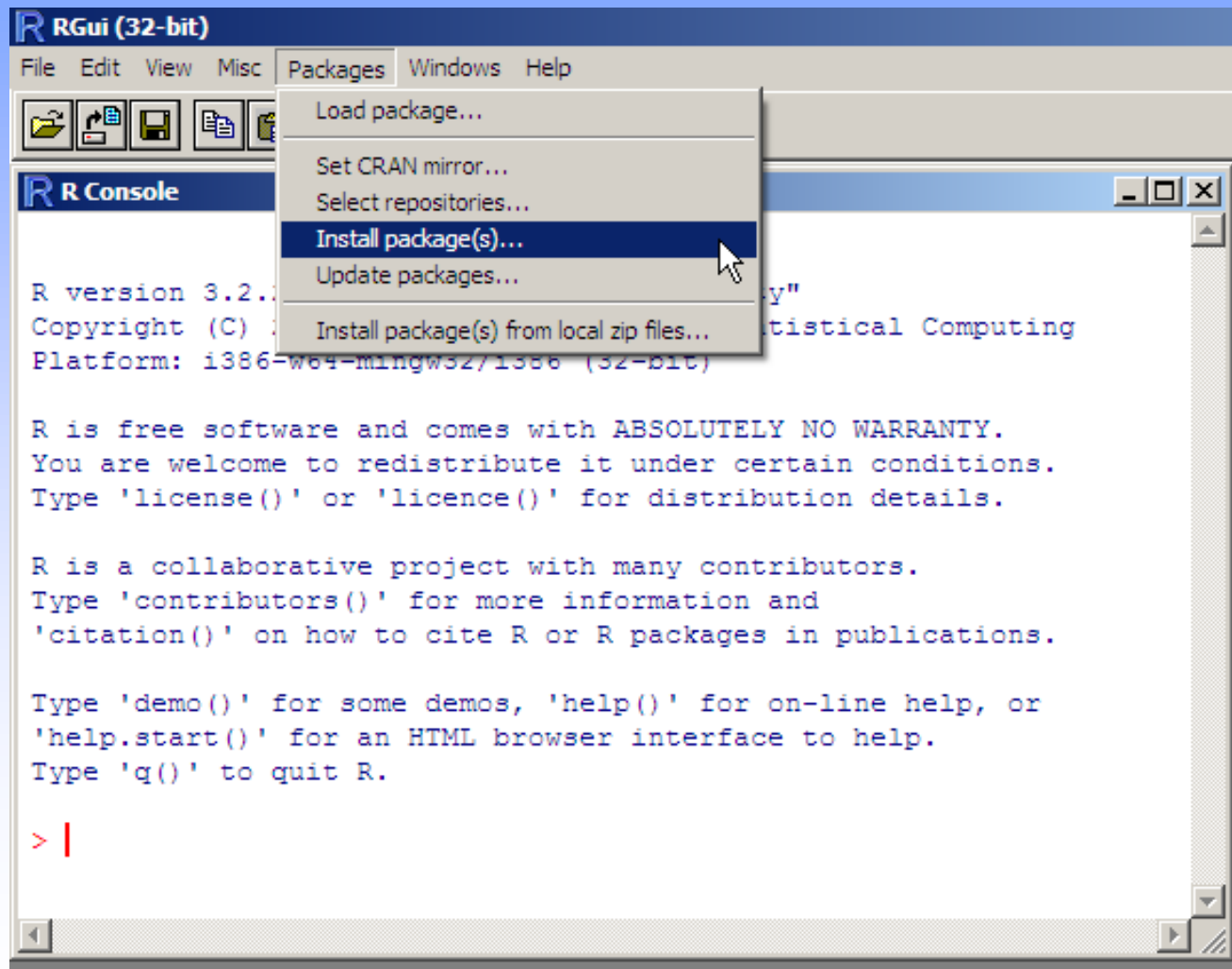
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

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.

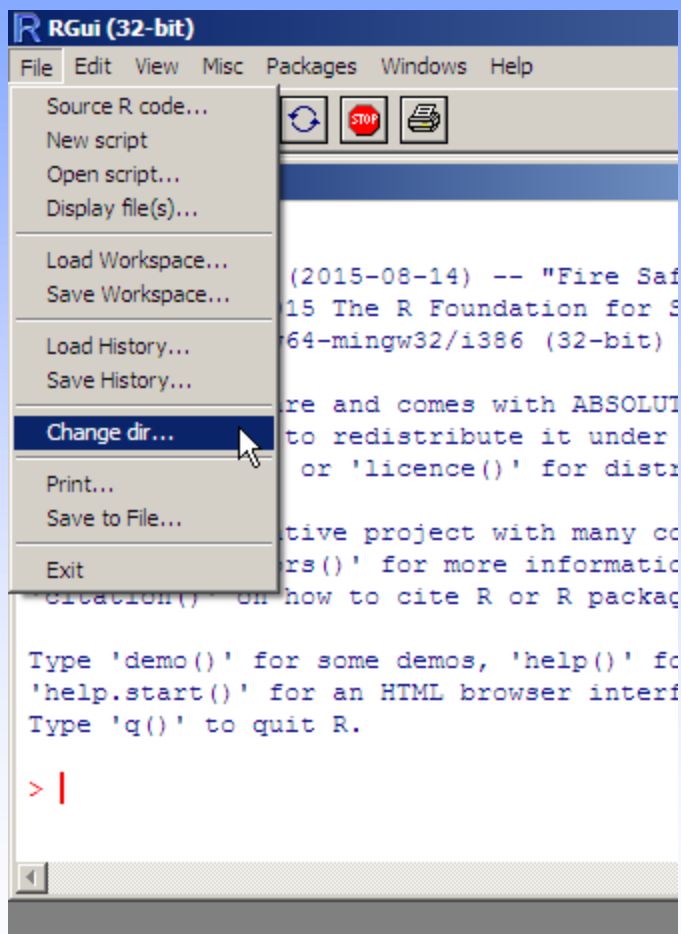
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> |
```

R pakešu instalācija



Darba sākums R



Norāda
noklusēto datu
direktoriju

Palīgi darbā ar R

- <https://cran.r-project.org/manuals.html>
- www.r-tutor.com
- dendro.daba.lv/R

Komandrinda

- Pārsvarā visas darbības veic, izpildot komandas
 - komandas var rakstīt aiz > zīmes un nospieš **Enter** taustiņu
 - komandas var rakstīt teksta redaktorā un pēc tam pārkopēt
 - komandas var saglabāt teksta failā ar paplašinājumu *.R un izpildīt

Komandas

- Var izmantot kā kalkulatoru

> 3 + 5

[1] 8

> 2/4

[1] 0.5

> 2 ^ 2

[1] 4

Mainīgie

- Vērtības var piešķirt mainīgajiem un darbības veikt ar tiem

```
> x <- 2
```

```
> y = 4
```

```
> z = x + y
```

```
> z
```

```
[1] 6
```

Datu veidi

- Skaitliskie (numeric)
5.17, 18.5
- Veselie skaitļi (integer)
3, 5
- Zīmju (character)
Jānis, Pēteris
- Loģiskie (logical)
TRUE, FALSE

Vērtību piešķiršana mainīgajiem

```
> a = 3
```

```
> a
```

```
[1] 3
```

```
> b = "Jānis"
```

```
> b
```

```
[1] "Jānis"
```

```
> c = TRUE; c ; d = FALSE ; d
```

```
[1] TRUE
```

```
[1] FALSE
```

Datu struktūras

- Vektori (vector)
- Matrices (matrix)
- Saraksts (list)
- Datu tabula (data frame)

Vektori

- Viendimensionāla struktūra, kas sastāv no viena veida datiem (skaitliskiem, zīmju, loģiskiem)
- Definējot vektoru jāraksta = vai <- atbilstošā vērtība

```
> a <- 5
```

```
> a
```

```
[1] 5
```

Vektori

- Vektors var ietvert vienu vai neierobežotu skaitu elementu
- Lai vektorā ietvertu vairākus mainīgos, jāizmanto funkcija `c()`

```
> a = c (1,2,3,5) ; a
```

```
[1] 1 2 3 5
```

```
> b = c ("a", "b", "c", "e")
```

```
[1] "a", "b", "c", "e"
```

Vektoru apvienošana

- Vektorus veidojot, var izmantot ne tikai vērtības, bet arī citus vektorus

> c = c (9, 11, 18)

> d = c (a, c)

> d

[1] 1 2 3 5 9 11 18

Regulāru vektoru veidošana

- Funkcija `seq()`

```
> v1 = seq(from = 1, to = 7, by = 1)
```

```
> v1
```

```
[1] 1 2 3 4 5 6 7
```

Regulāru vektoru veidošana

- Funkcija `rep()`

```
> v2 = rep(x = c("A", "B", "C"), times = 2)
```

```
> v2
```

```
[1] "A" "B" "C" "A" "B" "C"
```

```
> v3 = rep(x = c("A", "B", "C"), each = 2)
```

```
> v3
```

```
[1] "A" "A" "B" "B" "C" "C"
```

Darbības ar vektoriem

Vektorus var izmantot, lai veiktu aprēķinus ar tajā iekļautajiem skaitļiem, piemēram, aprēķinātu **vidējo** (*mean()*), **maksimālo** (*max()*) **vērtību**, **mediānu** (*median()*), **diapazonu** (*range()*), **elementu summu** (*sum()*) u.c.

Aritmētiskās darbības

- Pieskaitot vektoram vienu skaitli, tas pieskaitās visiem elementiem
- Saskaitot divus vektorus (vienāda garuma), sumējas to atsevišķie elementi

Logiskie operatori

- d

[1] 1 2 3 5 9 11 18

- lg1 = d == 5 ; lg1

– [1] FALSE FALSE FALSE TRUE FALSE FALSE FALSE

- lg2 = d >= 9 ; lg2

– FALSE FALSE FALSE FALSE TRUE TRUE TRUE

Vektora elementu izvēle

```
> d[4]
```

```
[1] 5
```

```
> d[c(1,3,5,7)]
```

```
[1] 1 3 9 18
```

```
> d[lg1]
```

```
[1] 5
```

```
> d[d>6]
```

```
[1] 9 11 18
```

Vektora elementu izvēle

```
> d[2:5]
```

```
[1] 2 3 5 9
```

```
> v4 = c(v3,"A"); v4
```

```
"A" "A" "B" "B" "C" "C" "A"
```

```
> lg4 = v4 == "A" ; lg4
```

```
TRUE TRUE FALSE FALSE FALSE FALSE TRUE
```

```
> d[lg4]
```

```
[1] 1 2 18
```

Matrices (Matrix)

- Divdimensionāls masīvs ar viena veida datiem
 - Var veidot, apvienojot vienāda garuma vektorus vai ar funkciju **matrix()**
- ```
> mt1 = cbind(d,f,g)
> mt2 = rbind(d,f,g)
> mt3 = matrix(data = c(1,2,3,4,5,6), nrow = 3, ncol = 2)
```

# Saraksts (list)

- Tajā var apvienot atšķirīgas datu struktūras, piemēram, vektorus un matrices
  - Bieži izmanto statistisko analīžu rezultātu saglabāšanai
  - Veido ar funkciju **list()**
- > ls = list(vektors = v4, matrice = mt2)

# Datu tabula (Data frame)

- Datu tabula, kuras katrā ailē var būt dažāda veida dati (skaitliski, zīmju, loģiskie)
  - Veido ar funkciju **data.frame()** vai iegūst, ielasot datus no faila
- ```
> tab1 = data.frame(paraug1 = v4, dati = mt1)
```
- Tabulas labošana:
- ```
> edit(tab1)
```

# Datu importēšana

## Teksta failu ielasīšana

```
> tab2a = read.table(file = "piemers_1a.txt",
header = TRUE, sep = "\t", dec = ".")
```

- file – ielasāmais fails
- header – vai pirmā rindā ir aiļu nosaukumi
- sep – atdalītājs
- dec – decimālā zīme

# Datu importēšana

Pirmo aili ielasa kā rindu nosaukumus

```
> tab2b = read.table(file = "piemers_1b.txt",
 header = TRUE, sep = "\t", dec = ".")
```

CSV formāta failu ielasīšana

```
> tab3 = read.csv(file = "piemers_1a.csv",
 header = TRUE, sep = ";", dec = ".")
```

# Rindu un aiļu nosaukumi

Aiļu nosaukumu apskate

```
> names(tab2a)
```

Rindu nosaukumu apskate

```
> row.names(tab2b)
```

# Aiņu izvēle

> tab2a\$gads

> tab2a[ ,1]

> tab2a\$gpl

> tab2a[ ,2]

> tab2a[ ,3:4]

# Aktīvie objekti

Ar funkciju `ls()` var noskaidrot aktīvos objektus šajā sesijā

> `ls()`

# Darba sesijas saglabāšana

- File – Save History ...
  - saglabā visas izpildītās komandas failā \*.Rhistory
- File – Save Workspace ...
  - saglabā visus objektus failā \*.Rdata
  - tas nākamreiz atverams ar File – Load Workspace
- File – Save to File ...
  - saglabā visas komandas un rezultātus teksta failā

# Grafiki

Funkcija **plot()** zīmē izkliedes grafiku (*scatterplot*), ja norāda vienu vai divus skaitlisku datu vektorus, vai *box-plot* grafiku, ja viens vektors nav skaitlisks

```
> plot(tab2a$gpl)
```

```
> plot(tab2a$gads, tab2a$gpl)
```

Var norādīt citu grafika veidu

```
> plot(tab2a$gads, tab2a$gpl, type = "l")
```

# Grafiku veidi

p – punkti

l – līnija

b – abi

h – histogrammai līdzīgas vertikālas līnijas

n – grafiku neveido (paliek tukšas asis)

# Histogramma

Funkcija **hist()** veido histogrammu no viena skaitlisku datu vektora

```
> hist(tab2a$gpl)
```

Funkcija **boxplot()** veido *Box-plot* grafiku vienam vai vairākiem datu vektoriem

```
> boxplot(tab2a$gpl)
```

```
> boxplot(tab2a[,3:4])
```