





ساخت، فراخوانی، تغییر داده ها در R

مدرس: مریم عبدلی
دانشجوی دکتری آمارزیستی

فهرست

- مقدمه
- خواندن داده ها از سایر نرم افزار ها
- ذخیره داده ها در فرمت فایل متنی، اکسل، spss
- فراخوانی داده های موجود در بسته های R
- ساخت داده:ذخیره، فراخوانی و ویرایش بردار
- ساخت داده:ذخیره، فراخوانی و ویرایش ماتریس
- ساخت داده:ذخیره، فراخوانی و ویرایش چارچوب اطلاعاتی

مقدمه

- داده ها نوعی شی هستند که می توان آن ها را با شیوه های مختلف وارد نرم افزار R کرد.

- سه روش ورود این اشیاء :

1. ورود مستقیم از طریق صفحه کلید

2. خواندن داده های ذخیره شده با فرمت های دیگر

3. دریافت داده های ذخیره شده از اینترنت

ساخت بردار

برای ساختن بردار 4 تابع عمده وجود دارد که عبارتند از:

`C()`

تابع `scan()`

تابع `rep()`

تابع `seq()`

```
> x<-c(2,5,10,11,30)
> x
[1] 2 5 10 11 30
> x1<-scan()
1: 12
2: 13
3: 16
4: 20
5:
Read 4 items
> rep(2, times = 5)
[1] 2 2 2 2 2
> rep(c(2, 5, 10), times = 2)
[1] 2 5 10 2 5 10
> rep(c(2, 10), times = c(2,3))
[1] 2 2 10 10 10
> rep(c(2,5), each=2)
[1] 2 2 5 5
> seq(-1,2,by=0.5)
[1] -1.0 -0.5 0.0 0.5 1.0 1.5 2.0
> |
```

ساخت ماتریس

برای ساختن ماتریس 4 تابع وجود دارد که عبارتند از:

`matrix()`

تابع `cbind()`

تابع `rbind()`

تابع `diag()`

ساخت ماتریس

```
> matrix(data=c(4,5,7,6,9,3),nrow=2,byrow=FALSE)
     [,1] [,2] [,3]
[1,]    4    7    9
[2,]    5    6    3
> matrix(data=c(4,5,7,6,9,3),nrow=2,byrow=TRUE)
     [,1] [,2] [,3]
[1,]    4    5    7
[2,]    6    9    3
> x1<-1:4
> x2<-5:8
> x3<-9:12
> x<-cbind(x1,x2,x3)
> x
      x1 x2 x3
[1,]  1  5  9
[2,]  2  6 10
[3,]  3  7 11
[4,]  4  8 12
> x1<-matrix(data=1:12,nrow=2,byrow=TRUE)
> x1<-matrix(data=1:12,nrow=2,byrow=TRUE)
> x2<-matrix(data=20:23,nrow=2,byrow=FALSE)
> cbind(x1,x2)
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
[1,]    1    2    3    4    5    6   20   22
[2,]    7    8    9   10   11   12   21   23
> x1<-1:4
> x2<-5:8
> x3<-9:12
> x<-rbind(x1,x2,x3)
> x
     [,1] [,2] [,3] [,4]
x1    1    2    3    4
x2    5    6    7    8
x3    9   10   11   12
```

```
> diag(c(1,2,2,4))
     [,1] [,2] [,3] [,4]
[1,]    1    0    0    0
[2,]    0    2    0    0
[3,]    0    0    2    0
[4,]    0    0    0    4
> |
```

خواندن داده ها از سایر نرم افزار ها

- R دارای توابع زیادی است که این نرم افزار را قادر می سازد با نرم افزارها و پایگاه های داده ، تبادل اطلاعات کند. به طور مثال توابع
 - `read.table()` برای خواندن داده های فایل متنی
 - `read.csv()` برای خواندن داده های فایل Excel
 - `read.SPSS()` برای خواندن داده های فایل spss
 - `read.ssd()` برای خواندن داده های فایل SAS
 - `read.mtp()` برای خواندن داده های فایل Minitab
 - `Read.dta()` برای خواندن داده های فایل STATA می باشند.

خواندن داده ها از سایر نرم افزار ها

- `read.table()` و `read.csv()` این دو تابع به صورت خودکار فراخوانی میشوند.
- بقیه توابع در بسته `foreign` گنجانده شده اند. برای استفاده از آن ها باید این بسته فراخوانی شود.

خواندن فایل متنی

- شکل کلی و خلاصه شده تابع `read.table()`

`read.table(file, header, sep)`

آرگومان `file`: آدرس فایل متنی که بایستی درون " " محصور شود

آرگومان `header`: عبارت منطقی که بودن یا نبودن نام متغیرها در سطر اول فایل را مشخص می کند

آرگومان `sep`: کاراکتر جداکننده مقادیر هر سطر فایل که درون "" معرفی شده و مشخص می کند.

خواندن فایل و تغییر داده

```
> data <- read.table("C:\\Users\\Hooshmand\\Desktop\\data.txt", header = TRUE, sep = "")
> data
  neзад pass frequencu sex
1      2    1         5   1
2      2    2         1   2
3      1    1         4   1
4      1    2         1   2
> # اضافه کردن لیبلها به ستون neзад
> data$neзад <- factor(data$neзад, levels = c(1, 2), labels = c("siah", "sefid"))
> data
  neзад pass frequencu sex
1 sefid    1         5   1
2 sefid    2         1   2
3 siah     1         4   1
4 siah     2         1   2
> data$pass <- factor(data$pass, levels = c(1, 2), labels = c("yes", "no"))
> data
  neзад pass frequencu sex
1 sefid yes         5   1
2 sefid no          1   2
3 siah yes         4   1
4 siah no          1   2
> data$sex <- factor(data$sex, levels = c(1, 2), labels = c("girl", "boy"))
> data
  neзад pass frequencu sex
1 sefid yes         5 girl
2 sefid no          1 boy
3 siah yes         4 girl
4 siah no          1 boy
```

خواندن فایل Excel

- برای خواندن یک فایل داده از Excel، می توان از تابع `read.csv` یا `read.table()` استفاده کرد.
- البته برای استفاده از این توابع، باید ابتدا فایل Excel را در قالب یک فایل متنی (فرمت CSV . و یا TXT) ذخیره کرد.

خواندن فایل Excel

- برای ذخیره در قالب CSV. سه گزینه متفاوت ذخیره سازی در Excel تعبیه شده است
- برای ذخیره در قالب TXT. نیز سه گزینه متفاوت ذخیره سازی در Excel تعبیه شده است. عبارتند از Text(Macintosh)، Text(Tab delimited) و Text(MS-DOS).
- برای فراخوانی فایل ذخیره شده در قالب اول از `sep="\t"` و در دو مورد دیگر از `sep=""` استفاده میشود.

خواندن فایل اکسل و تغییر داده

```
> dataset <- read.table("C:\\Users\\Hooshmand\\Desktop\\data.csv", header=TRUE, sep=",")
> dataset
  nezap pass frequencu
1     2    1         5
2     2    2         1
3     1    1         4
4     1    2         1
>
> # اضافه کردن لیبلها به ستون nezap
> dataset$nezap <- factor(dataset$nezap, levels = c(1, 2), labels = c("siah", "sefid"))
> dataset
  nezap pass frequencu
1 sefid    1         5
2 sefid    2         1
3  siah    1         4
4  siah    2         1
> dataset$pass <- factor(dataset$pass, levels = c(1, 2), labels = c("yes", "no"))
> dataset
  nezap pass frequencu
1 sefid  yes         5
2 sefid  no          1
3  siah  yes         4
4  siah  no          1
```

خواندن فایل SPSS

- برای خواندن داده های SPSS باید از تابع `read.spss()` در بسته `foreign` استفاده کرد. شکل کلی و خلاصه شده تابع `read.spss()` به صورت زیر است

`read.spss(file, to.data.frame)`

آرگومان `file`: آدرس فایل متنی است که بایستی درون `""` محصور شود

آرگومان `to.data.frame`: عبارتی منطقی است که خواندن در قالب چارچوب اطلاعاتی یا لیست را مشخص می کند

خواندن فایل SPSS

```
> # First, install and load the foreign package if you haven't already
> install.packages("foreign")
Warning: package 'foreign' is in use and will not be installed
> library(foreign)
> # Now, use the read.spss function to load your .sav file
> mydata <- read.spss("C:\\Users\\Hooshmand\\Desktop\\Untitled2.sav", to.data.frame=TRUE)
> mydata
  neza pass frequen
1 sefid  yes      5
2 sefid  no       1
3 siah   yes      4
4 siah   no       1
> |
```

ذخیره داده ها در قالب فایل متنی (Export)

- تقریباً تمامی نرم افزار های تحلیل داده امکان خواندن داده های ذخیره شده در فایل متنی را دارند. از این رو برای استفاده از داده های R در دیگر نرم افزار ها، کافی است آن ها را در قالب یک فایل متنی ذخیره کرد. به این منظور می توان از تابع `write.table()` استفاده کرد.

ذخیره داده ها در قالب فایل متنی (Export)

- شکل کلی و خلاصه شده این تابع عبارت است از:

write.table(x, file)

آرگومان X: بردار، ماتریس، چارچوب اطلاعاتی یا هر شی دیگری که قصد ذخیره کردن آن را داریم

آرگومان file: آدرس محل ذخیره سازی و نام فایل متنی است که بایستی درون ""محصور شود

ذخیره داده ها در قالب فایل متنی (Export)

برای نمایش و ذخیره کردن چارچوب اطلاعاتی mydata در قالب فایل متنی Rxdata.txt در مسیر دسکتاب، به صورتی که مقادیر یک سطر با فاصله از هم جدا شوند،

```
> mydata
  nezad pass frequencu
1 sefid  yes         5
2 sefid  no          1
3 siah   yes         4
4 siah   no          1
> write.table(x=mydata, "C:\\Users\\Hooshmand\\Desktop\\Rxdata.txt")
```

ذخیره داده ها در قالب فایل اکسل (Export)

```
> if (!require(writexl)) {  
+   install.packages("writexl")  
+   library(writexl)  
+ }  
>  
> # ذخیره داده‌ها در یک فایل اکسل  
> write_excel(dataset, path = "C:\\\\Users\\Hooshmand\\Desktop\\data.xlsx")  
> |
```

ذخیره داده ها در قالب فایل spss (Export)

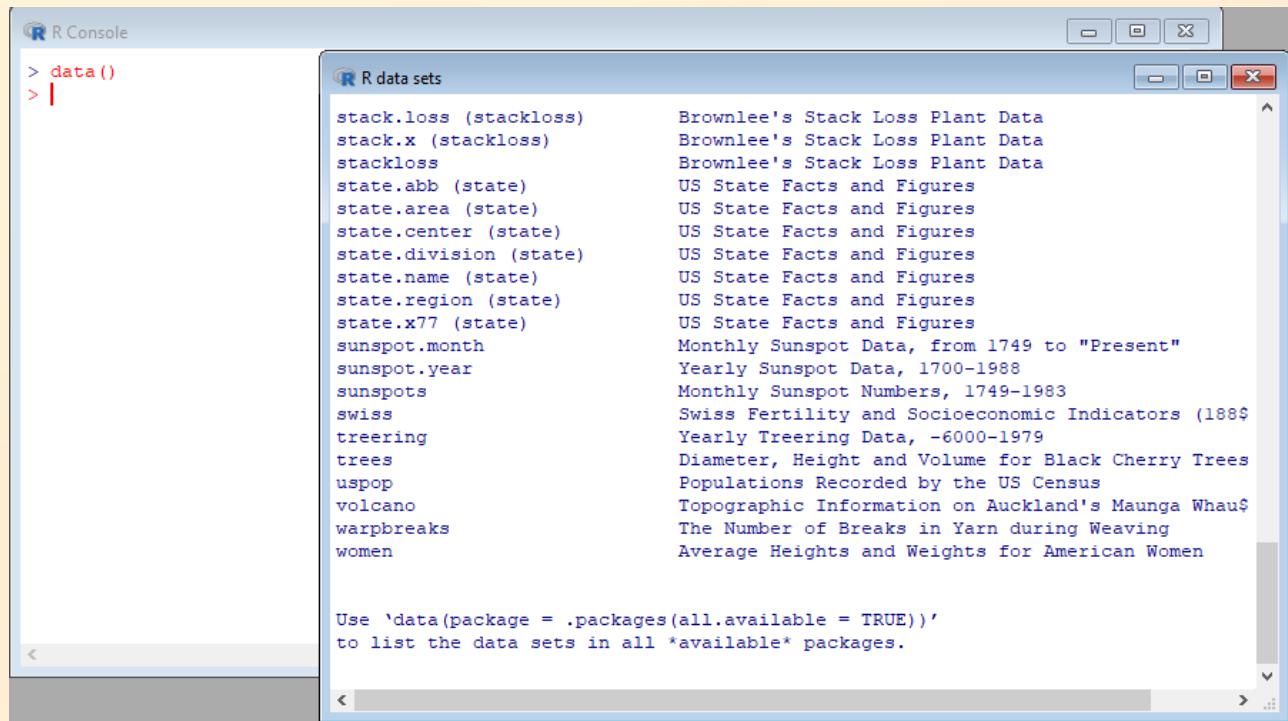
```
> if (!require(haven)) {  
+   install.packages("haven")  
+   library(haven)  
+ }  
>  
> # ذخیره داده ها در فایل SPSS  
> write_sav(dataset, path = "C:\\Users\\Hooshmand\\Desktop\\data.sav")  
> |
```

تمرین: جدول زیر را در R ایجاد کنید و سپس با اسم A1 در مسیر دسکتاپ با فرمت فایل متنی ذخیره کنید

name	gender	age	test
ezat	m	20	T
efat	f	15	T
goli	f	17	F
heshmat	m	22	F

فراخوانی داده های موجود در بسته های R

مجموعه ای از داده های واقعی برای مقاصد آموزشی و تمرینی در بسته نرم افزاری datasets گنجانده شده است. برای مشاهده فهرست این داده ها می توان از کد زیر استفاده کرد



```
R Console
> data()
> |

R data sets
stack.loss (stackloss)      Brownlee's Stack Loss Plant Data
stack.x (stackloss)        Brownlee's Stack Loss Plant Data
stackloss                  Brownlee's Stack Loss Plant Data
state.abb (state)          US State Facts and Figures
state.area (state)         US State Facts and Figures
state.center (state)       US State Facts and Figures
state.division (state)     US State Facts and Figures
state.name (state)         US State Facts and Figures
state.region (state)       US State Facts and Figures
state.x77 (state)          US State Facts and Figures
sunspot.month              Monthly Sunspot Data, from 1749 to "Present"
sunspot.year               Yearly Sunspot Data, 1700-1988
sunspots                   Monthly Sunspot Numbers, 1749-1983
swiss                      Swiss Fertility and Socioeconomic Indicators (1880-1985)
treering                   Yearly Treering Data, -6000-1979
trees                      Diameter, Height and Volume for Black Cherry Trees
uspop                      Populations Recorded by the US Census
volcano                    Topographic Information on Auckland's Maunga Whau
warpbreaks                 The Number of Breaks in Yarn during Weaving
women                      Average Heights and Weights for American Women

Use 'data(package = .packages(all.available = TRUE))'
to list the data sets in all *available* packages.
```

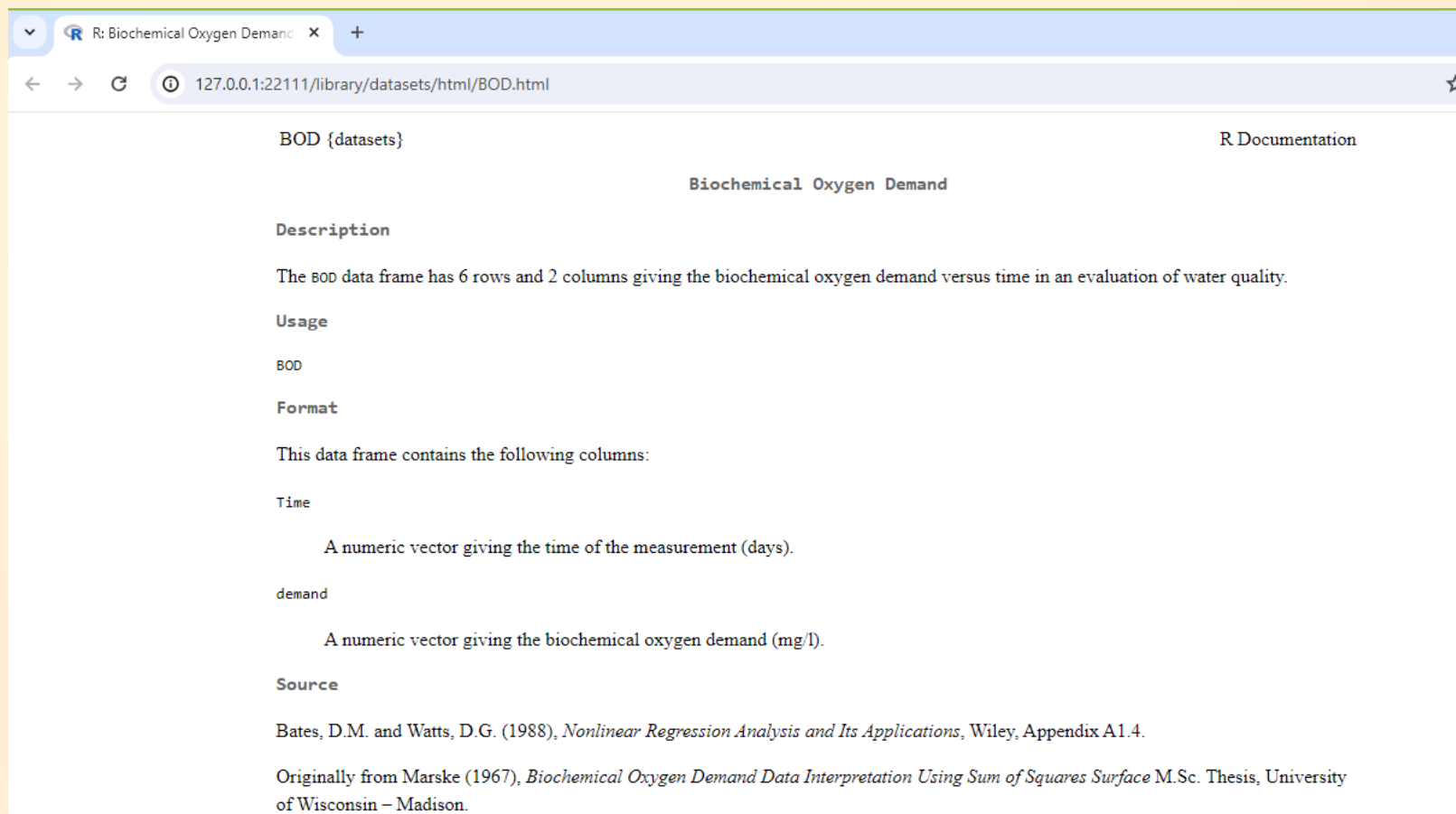
فراخوانی داده های موجود در بسته های R

برای دیدن جزئیات کامل تری از هر داده، میتوان به صفحه راهنمای آن مراجعه کرد.

برای نمونه، جزئیات داده BOD در این بسته را می توان با استفاده از کد برنامه زیر استخراج کرد.

```
> data()
> help(BOD)
starting httpd help server ... done
> |
```

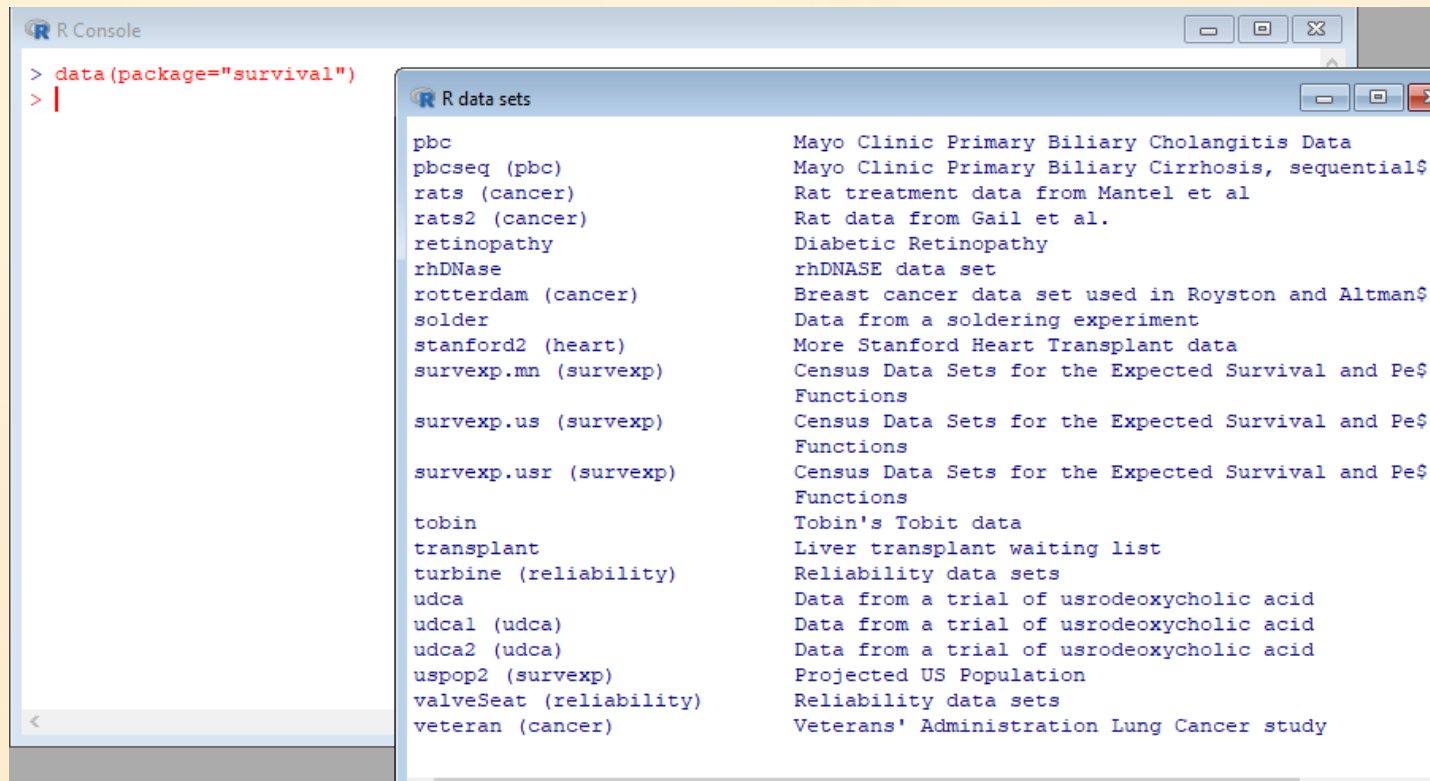
فراخوانی داده های موجود در بسته های R



The screenshot shows a web browser window displaying the R documentation for the 'BOD' dataset. The browser's address bar shows the URL '127.0.0.1:22111/library/datasets/html/BOD.html'. The page title is 'BOD {datasets}' and 'R Documentation'. The main heading is 'Biochemical Oxygen Demand'. The 'Description' section states: 'The BOD data frame has 6 rows and 2 columns giving the biochemical oxygen demand versus time in an evaluation of water quality.' The 'Usage' section shows the command 'BOD'. The 'Format' section states: 'This data frame contains the following columns:'. The 'Time' section describes the 'time' column as 'A numeric vector giving the time of the measurement (days)'. The 'demand' section describes the 'demand' column as 'A numeric vector giving the biochemical oxygen demand (mg/l)'. The 'Source' section lists 'Bates, D.M. and Watts, D.G. (1988), *Nonlinear Regression Analysis and Its Applications*, Wiley, Appendix A1.4.' and 'Originally from Marske (1967), *Biochemical Oxygen Demand Data Interpretation Using Sum of Squares Surface* M.Sc. Thesis, University of Wisconsin – Madison.'

فراخوانی داده های موجود در بسته های R

- سایر بسته های نرم افزاری R ممکن است علاوه بر توابع، شامل تعدادی مجموعه داده باشند. برای مشاهده مجموعه داده هر بسته نرم افزاری، مانند Survival، میتوان از کد برنامه زیر استفاده کرد.



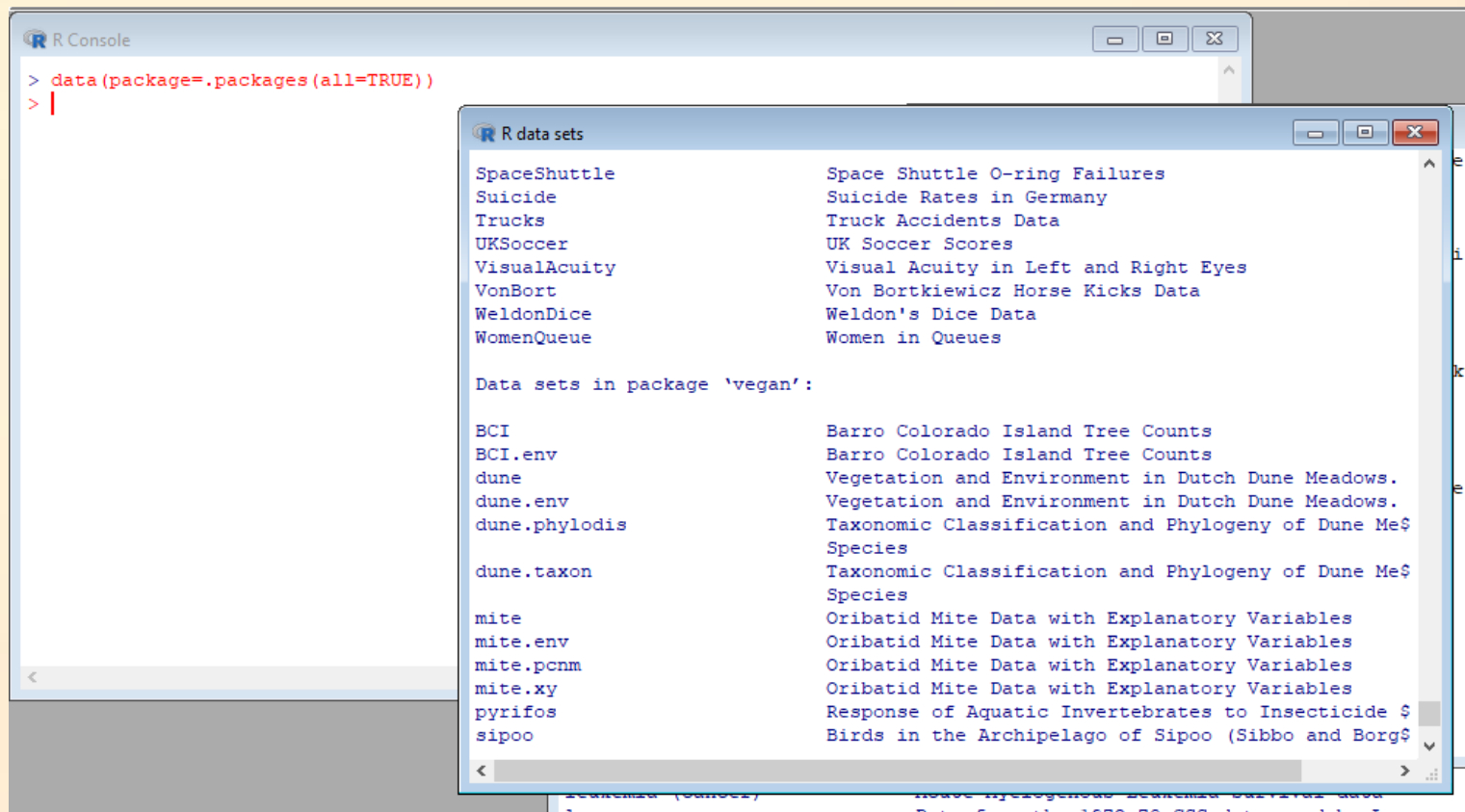
```
R Console
> data(package="survival")
> |

R data sets
pbcb          Mayo Clinic Primary Biliary Cholangitis Data
pbcbseq (pbcb) Mayo Clinic Primary Biliary Cirrhosis, sequential$
rats (cancer)  Rat treatment data from Mantel et al
rats2 (cancer) Rat data from Gail et al.
retinopathy    Diabetic Retinopathy
rhDNase        rhDNASE data set
rotterdam (cancer) Breast cancer data set used in Royston and Altman$
solder         Data from a soldering experiment
stanford2 (heart) More Stanford Heart Transplant data
survexp.mn (survexp) Census Data Sets for the Expected Survival and Pe$
                Functions
survexp.us (survexp) Census Data Sets for the Expected Survival and Pe$
                Functions
survexp.usr (survexp) Census Data Sets for the Expected Survival and Pe$
                Functions
tobin          Tobin's Tobit data
transplant     Liver transplant waiting list
turbine (reliability) Reliability data sets
udca           Data from a trial of usrodeoxycholic acid
udca1 (udca)   Data from a trial of usrodeoxycholic acid
udca2 (udca)   Data from a trial of usrodeoxycholic acid
uspop2 (survexp) Projected US Population
valveSeat (reliability) Reliability data sets
veteran (cancer) Veterans' Administration Lung Cancer study
```

فراخوانی داده های موجود در بسته های R

- برای فراخوانی داده از بسته های نرم افزاری فراخوانی نشده، ابتدا باید آن بسته نرم افزاری فراخوانی کرد. همچنین برای مشاهده فهرست تمام مجموعه داده های ذخیره شده در همه بسته های نرم افزاری نصب شده روی سیستم، می توان از کد برنامه زیر استفاده کرد.

فراخوانی داده های موجود در بسته های R



```
> data(package=.packages(all=TRUE))
> |
```

R data sets

SpaceShuttle	Space Shuttle O-ring Failures
Suicide	Suicide Rates in Germany
Trucks	Truck Accidents Data
UKSoccer	UK Soccer Scores
VisualAcuity	Visual Acuity in Left and Right Eyes
VonBort	Von Bortkiewicz Horse Kicks Data
WeldonDice	Weldon's Dice Data
WomenQueue	Women in Queues

Data sets in package 'vegan':

BCI	Barro Colorado Island Tree Counts
BCI.env	Barro Colorado Island Tree Counts
dune	Vegetation and Environment in Dutch Dune Meadows.
dune.env	Vegetation and Environment in Dutch Dune Meadows.
dune.phylodis	Taxonomic Classification and Phylogeny of Dune Me\$ Species
dune.taxon	Taxonomic Classification and Phylogeny of Dune Me\$ Species
mite	Oribatid Mite Data with Explanatory Variables
mite.env	Oribatid Mite Data with Explanatory Variables
mite.pcnm	Oribatid Mite Data with Explanatory Variables
mite.xy	Oribatid Mite Data with Explanatory Variables
pyrifos	Response of Aquatic Invertebrates to Insecticide \$
sipoo	Birds in the Archipelago of Sipoo (Sibbo and Borg\$

ساخت داده: ذخیره، فراخوانی و ویرایش بردار

- مثال زیر یک بردار به نام `x` ذخیره و فراخوانی می شود

```
> x<-seq(-2,2,length.out=5)
> x
[1] -2 -1  0  1  2
> x<-seq(-2,2,length.out=4)
> x
[1] -2.0000000 -0.6666667  0.6666667  2.0000000
> x<-seq(-2,2,length.out=5)
> x
```

ساخت داده:ذخیره، فراخوانی و ویرایش بردار

- برای انتخاب یک قسمت از بردار چندین روش وجود دارد که در ادامه به ۲ روش آن اشاره میشود

- **روش اول:** در این روش میتوان شماره (اندیس)درايه های که باید انتخاب شوند

```
> vec<-c(7,2,6,3,12)
> vec[3]
[1] 6
> vec[c(2,3)]
[1] 2 6
> |
```

داخل “[]”قرار داد

ساخت داده: ذخیره، فراخوانی و ویرایش بردار

- **روش دوم:** این روش زمانی مناسب است که درایه های مورد نظر دارای ویژگی خاصی باشند. به عنوان نمونه در کد برنامه های زیر درایه های X بزرگتر از ۴۰ انتخاب شده اند.

```
> x1<-c(8,58,69,44)
> x1
[1] 8 58 69 44
> x1[x1>40]
[1] 58 69 44
> |
```

ساخت داده:ذخیره، فراخوانی و ویرایش ماتریس

- برای ذخیره سازی یک ماتریس بایستی به آن نامی اختصاص داد.برای فراخوانی یک ماتریس نام آن را در خط فرمان اجرا کرد

```
> x2<-matrix(c(4,5,7,6,9,1),nrow=3,byrow=FALSE)
> x2
      [,1] [,2]
[1,]    4    6
[2,]    5    9
[3,]    7    1
> |
```

ساخت داده: ذخیره، فراخوانی و ویرایش ماتریس

برای انتخاب قسمتی از ماتریس، می توان به محدوده سطر ها و ستون های حاوی قسمت مورد نظر اشاره کرد. این محدوده به وسیله “[r]” آدرس دهی میشود

مقدار یا مقادیری که (در قالب بردار) در **سمت راست** علامت “,” قرار میگیرند، **محدوده ستون ها**

و مقدار یا مقادیری که (در قالب بردار) در **سمت چپ** آن قرار می گیرند، **محدوده سطر ها** را مشخص می کند

ساخت داده: ذخیره، فراخوانی و ویرایش ماتریس

```
> x2<-matrix(c(4,5,7,6,9,1),nrow=3,byrow=FALSE)
> x2
      [,1] [,2]
[1,]    4    6
[2,]    5    9
[3,]    7    1
> x2[c(2,3),c(1,2)]
      [,1] [,2]
[1,]    5    9
[2,]    7    1
> x2[1,2]
[1] 6
> x2[2,c(1,2)]
[1] 5 9
> x2[-c(1,2),1]
[1] 7
```

ساخت داده: ذخیره، فراخوانی و ویرایش چارچوب اطلاعاتی

اطلاعات زیر در مورد 6 کودک که برای وضعیت آگاهی در زمینه بهداشت دست ها می باشد.

ID	Name.set	age	sex	hygine
1	Hamid	5	Boy	Bad
2	Ali	3	Boy	Ok
3	Arash	4	Boy	Ok
4	Babak	2	Boy	Good
5	Jalal	3	Boy	Ok
6	Amin	6	Boy	bad

ساخت داده: ذخیره، فراخوانی و ویرایش چارچوب اطلاعاتی

R Console

```
> ID<-factor(1:6)
> name.set<-c('hamid','ali','arash','babak','jalal','amin')
> age<-c(5,3,4,2,3,6)
> sex <- factor(rep(1, 6), levels = c(0, 1), labels = c("girl", "boy"))
> hygiene<-factor(c(3,2,2,1,2,3),levels=c(1,2,3),labels=c('good','ok','bad'),ordered=TRUE)
> data.frame(ID,name.set,age,sex,hygiene)
  ID name.set age sex hygiene
1  1   hamid   5 boy    bad
2  2    ali   3 boy    ok
3  3  arash   4 boy    ok
4  4  babak   2 boy    good
5  5  jalal   3 boy    ok
6  6   amin   6 boy    bad
> hyg.data<-data.frame(ID,name.set,age,sex,hygiene)
> |
```

ساخت داده: ذخیره، فراخوانی و ویرایش چارچوب اطلاعاتی

• گاهی کاربر مایل است قسمتی از یک چهارچوب داده را استخراج، ذخیره یا ویرایش کند. در اینصورت لازم است که کاربر قادر باشد به درایه های آن قسمت اشاره کند.

• **روش اول:** راه حل کلی استفاده از عملگر [,]

مقدار یا مقادیری که (در قالب بردار) در **سمت راست** علامت "," قرار میگیرند، **متغیرها**

و مقدار یا مقادیری که (در قالب بردار) در **سمت چپ** آن قرار می گیرند، **افراد مورد نظر** را

مشخص می کند

```
> hyg.data[3:5,2:4]
name.set age sex
3 arash 4 boy
4 babak 2 boy
5 jalal 3 boy
> |
```

ساخت داده: ذخیره، فراخوانی و ویرایش چارچوب اطلاعاتی

• روش دوم: از عملگر "\$" برای انتخاب(تنها) یک متغیر

هنگام استفاده از این عملگر، سمت چپ آن نام چارچوب اطلاعاتی و سمت راست آن نام متغیر مورد نظر قرار می گیرد

```
> hyg.data$name.set
[1] "hamid" "ali" "arash" "babak" "jalal" "amin"
> hyg.data$age
[1] 5 3 4 2 3 6
```

ساخت داده: ذخیره، فراخوانی و ویرایش چارچوب اطلاعاتی

• روش سوم: از عملگر "[]"

هنگامی استفاده می شود که کاربر مایل باشد یک یا چند متغیر را بر اساس شماره ستون های آن ها استخراج کند. در اینصورت شماره ستون متغیر های مورد نظر در قالب یک بردار عددی وارد می شود.

```
> hyg.data[2:4]
  name.set age sex
1   hamid   5 boy
2    ali   3 boy
3  arash   4 boy
4  babak   2 boy
5  jalal   3 boy
6   amin   6 boy
```



پایان

تشکر از توجه شما