



Courage
Inspiration
Trust
Youth
Uniqueness

SDSC3006 Lab

1-Introduction to R

Langming LIU langmiliu2-c@my.cityu.edu.hk

School of Data Science
City University of Hong Kong

Outline

1. Install R and Data packages
2. Basic commands
3. Preliminary data analysis

Install R and Data packages



[\[Home\]](#)

Download

[CRAN](#)

R Project

[About R](#)

[Contributors](#)

[What's New?](#)

[Mailing Lists](#)

[Bug Tracking](#)

[Conferences](#)

[Search](#)

R Foundation

[Foundation](#)

[Board](#)

[Members](#)

[Donors](#)

[Donate](#)

Documentation

[Manuals](#)

[FAQs](#)

[The R Journal](#)

[Books](#)

[Certification](#)

[Other](#)

Links

[Bioconductor](#)

[Related Projects](#)

The R Project for Statistical Computing

Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To [download R](#), please choose your preferred [CRAN mirror](#).

If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

News

- [The R Journal Volume 7/2](#) is available.
- [R version 3.2.3 \(Wooden Christmas-Tree\)](#) has been released on 2015-12-10.>
- [R version 3.1.3 \(Smooth Sidewalk\)](#) has been released on 2015-03-09.
- [useR! 2015](#), took place at the University of Aalborg, Denmark, June 30 - July 3, 2015.

Install R

R: The R Project for Statistical Computing



download R



choose a location



choose the system



install R for the first time



choose the version and follow instruction

Data Sets of R

- The ISLR2 package

(Data for an introduction to statistical learning with applications in R version)

- The MASS library

(Modern applied statistics with S)

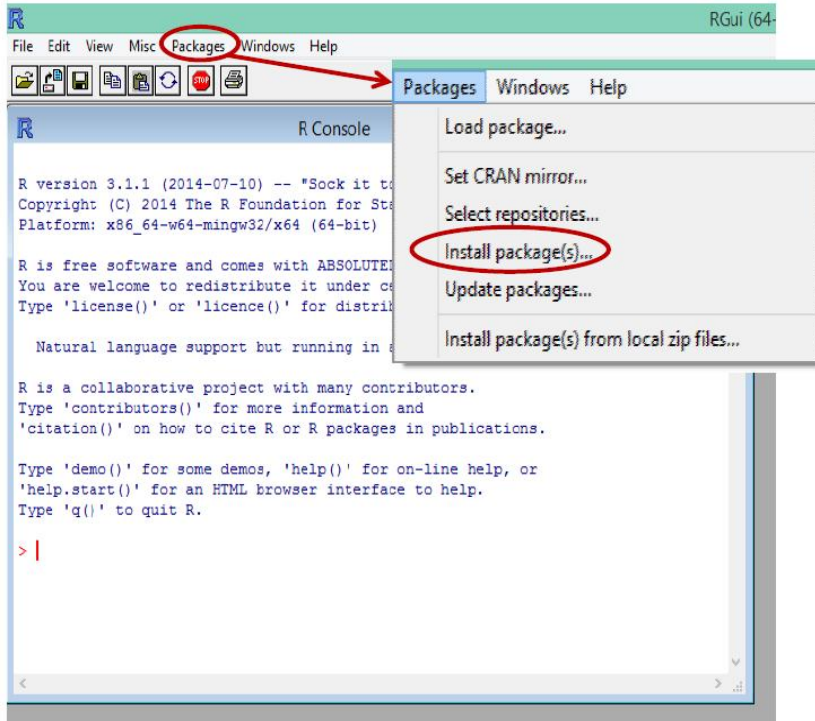
- Base R

Data Sets of R

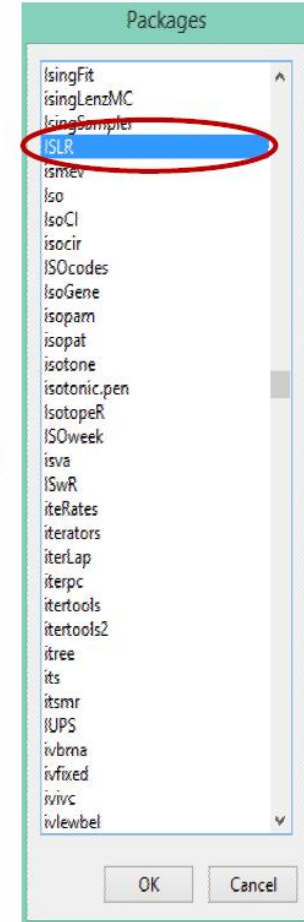
- The ISLR2 package

| Name | Description |
|------------------|---|
| Auto | Gas mileage, horsepower, and other information for cars. |
| Boston | Housing values and other information about Boston suburbs. |
| Caravan | Information about individuals offered caravan insurance. |
| Carseats | Information about car seat sales in 400 stores. |
| College | Demographic characteristics, tuition, and more for USA colleges. |
| Default | Customer default records for a credit card company. |
| Hitters | Records and salaries for baseball players. |
| Khan | Gene expression measurements for four cancer types. |
| NCI60 | Gene expression measurements for 64 cancer cell lines. |
| OJ | Sales information for Citrus Hill and Minute Maid orange juice. |
| Portfolio | Past values of financial assets, for use in portfolio allocation. |
| Smarket | Daily percentage returns for S&P 500 over a 5-year period. |
| USArrests | Crime statistics per 100,000 residents in 50 states of USA. |
| Wage | Income survey data for males in central Atlantic region of USA. |
| Weekly | 1,089 weekly stock market returns for 21 years. |

Install ISLR2 package



Choose CRAN mirror spot



Select **ISLR** package for installation

Or just type in console:
`install.packages('ISLR2')`

Basic commands

Vector

- Save things: “<-” or “=”
- Define a vector
 - use function c(): `x=c(1,2,3,4)`
 - use function seq() for arithmetic sequence:
`x=seq(from=a,to=b,by=c)` or `x=a:b`
- Find a component by index i: `a=x[i]`
- The length of vector: `length(x)`
- Delete vector: `rm(x)`

Vector

- Generate random vector:

`y=rnorm(20)` (standard normal)

`y=3*rnorm(20)+2` (non-standard normal)

`y=runif(n=20,min=-1,max=1)` (uniform)

Set Seed To reproduce the same random vector:

`set.seed(1)`

`rnorm(5)`

- Statistics calculation of vectors:

apply function “`sqrt()`”, “`mean()`”, “`var()`”, “`sd()`”, “`cor()`”, “`cov()`” and so on

Matrix

- Define a matrix using function “matrix()”:

```
A=matrix(1:20,nrow=5,ncol=4)
```

notice: default order of filling the matrix is by column
we can change the order:

```
A=matrix(1:20,nrow=5,ncol=4,byrow=TRUE)
```

- Some calculations of matrix is the same:

```
sqrt(A), mean(A), sum(A)
```

- dimension of matrix: `dim(A)`

Matrix

- Indexing of Matrix:

- obtain a component: $A[2,3]$
- obtain a row or column: “ $A[3,]$ ” for row, “ $A[,2]$ ” for column

- obtain a section:

$A[2:4,2:4]$

$A[c(2,4),c(2,4)]$ for selecting particular rows and columns

$A[-c(2,4),-c(2,4)]$ for deleting particular rows and columns

Preliminary data analysis

Load Dataset

- To load a data set in the ISLR2 package or other packages/libraries:

```
library(ISLR2)
```

- To load an external data set, first specify the directory:

“File” → “Change directory”

- If the data are saved as a **text file**:

```
Auto=read.table('Auto.data',header=T,na.strings='?')
```

- If the data are saved as a **csv file** (Excel):

```
stock=read.csv('0001.HK.csv',na.strings='?')
```

Basic operations of Dataset

- Type “names(datasetname)” to list all attributes (column names): names(stock)
- Dimension of Dataset: dim(stock)
- Access one attribute:
 - method 1: stock\$Low
 - method 2: “attach(stock)” then “Low”
- obtain numerical summaries: summary(stock)

Basic operations of Dataset

- generate figures of single variables using “plot()”:

```
attach(stock)
```

```
plot(High,type='l',xlab='date index',ylab='High  
price',main='Figure 1')
```

- generate plots of two variables:

```
attach(stock)
```

```
plot(Open,High,xlab='Open prcie',ylab='High  
price',main='Figure 2')
```