

Introduction to Data Science

Lab 1

Chapter 2 Lab: Introduction to R

Basic Commands

The screenshot displays the RStudio interface with the following components:

- Source Editor:** Contains R code for installing the ISLR package, creating vectors, matrices, and performing calculations.
- Console:** Shows the execution of the code, including the output of `ls()`, matrix creation, and calculations.
- Environment:** Shows the current environment with variables `x` and `num`.
- Packages:** Lists installed and available packages.

```
1 install.packages("ISLR")
2 library(ISLR)
3
4 # Chapter 2 Lab: Introduction to R
5
6 # Basic Commands
7
8 x <- c(1,3,2,5)
9 x
10 x = c(1,6,2)
11 x
12 y = c(1,4,3)
13 length(x)
14 length(y)
15 x + y
16 ls()
17 rm(x,y)
18 ls()
19
20 x <- matrix(data = c(1,2,3,4), nrow = 2, ncol = 2)
21 x
22 x <- matrix(c(1,2,3,4), 2, 2)
23 matrix(c(1,2,3,4), 2, 2, byrow = TRUE)
24 sqrt(x)
25 x^2
26
```

Console Output:

```
> ls()
character(0)
>
> x <- matrix(data = c(1,2,3,4), nrow = 2, ncol = 2)
> x
      [,1] [,2]
[1,]    1    3
[2,]    2    4
> x <- matrix(c(1,2,3,4), 2, 2)
> matrix(c(1,2,3,4), 2, 2, byrow = TRUE)
      [,1] [,2]
[1,]    1    2
[2,]    3    4
> sqrt(x)
      [,1] [,2]
[1,] 1.000000 1.732051
[2,] 1.414214 2.000000
> x^2
      [,1] [,2]
[1,]    1    9
[2,]    4   16
> |
```

Environment:

Variable	Value
x	num [1:2, 1:2] 1 2 3 4

Packages:

Name	Description	Version
User Library		
<input checked="" type="checkbox"/> ISLR	Data for an Introduction to Statistical Learning with Applications in R	1.2
System Library		
<input checked="" type="checkbox"/> base	The R Base Package	3.6.2
<input type="checkbox"/> boot	Bootstrap Functions (Originally by Angelo Canty for S)	1.3-23
<input type="checkbox"/> class	Functions for Classification	7.3-15
<input type="checkbox"/> cluster	"Finding Groups in Data": Cluster Analysis Extended Rousseeuw et al.	2.1.0
<input type="checkbox"/> codetools	Code Analysis Tools for R	0.2-16
<input type="checkbox"/> compiler	The R Compiler Package	3.6.2
<input checked="" type="checkbox"/> datasets	The R Datasets Package	3.6.2
<input type="checkbox"/> foreign	Read Data Stored by 'Minitab', 'S', 'SAS', 'SPSS', 'Stata', 'Systat', 'Weka', 'dBase', ...	0.8-72
<input checked="" type="checkbox"/> graphics	The R Graphics Package	3.6.2
<input checked="" type="checkbox"/> grDevices	The R Graphics Devices and Support for Colours and Fonts	3.6.2
<input type="checkbox"/> grid	The Grid Graphics Package	3.6.2
<input type="checkbox"/> KernSmooth	Functions for Kernel Smoothing Supporting Wand & Jones (1995)	2.23-16
<input type="checkbox"/> lattice	Trellis Graphics for R	0.20-38
<input type="checkbox"/> MASS	Support Functions and Datasets for Venables and Ripley's MASS	7.3-51.4
<input type="checkbox"/> Matrix	Sparse and Dense Matrix Classes and Methods	1.2-18
<input checked="" type="checkbox"/> methods	Formal Methods and Classes	3.6.2
<input type="checkbox"/> mgcv	Mixed GAM Computation Vehicle with Automatic Smoothness Estimation	1.8-31
<input type="checkbox"/> nlme	Linear and Nonlinear Mixed Effects Models	3.1-142
<input type="checkbox"/> nnet	Feed-Forward Neural Networks and Multinomial Log-Linear Models	7.3-12
<input type="checkbox"/> parallel	Support for Parallel computation in R	3.6.2
<input type="checkbox"/> rpart	Recursive Partitioning and Regression Trees	4.1-15

Simulation

The screenshot displays the RStudio interface with a script editor, console, and package manager.

Script Editor:

```
1 ## simulation
2 # set.seed: Set the seed of R's random number generator,
3 # which is useful for creating simulations or random objects that can be reproduced.
4 set.seed(1303)
5 rnorm(5)
6
7 set.seed(1303)
8 rnorm(5)
9
10 set.seed(3)
11 y <- rnorm(100)
12 mean(y)
13 var(y)
14 sqrt(var(y))
15 sd(y)
16
```

Console:

```
> ## simulation
> # set.seed: Set the seed of R's random number generator,
> # which is useful for creating simulations or random objects that can be reproduced.
> set.seed(1303)
> rnorm(5)
[1] -1.14397631  1.34212937  2.18539048  0.53639252  0.06319297
>
> set.seed(1303)
> rnorm(5)
[1] -1.14397631  1.34212937  2.18539048  0.53639252  0.06319297
>
> set.seed(3)
> y <- rnorm(100)
> mean(y)
[1] 0.01103557
> var(y)
[1] 0.7328675
> sqrt(var(y))
[1] 0.8560768
> sd(y)
[1] 0.8560768
```

Environment:

Object	Class	Value
x	num	[1:2, 1:2] 1 2 3 4
y	num	[1:100] -0.962 -0.293 0.259 -1.152 0.196 ...

Packages:

Name	Description	Version
User Library		
<input checked="" type="checkbox"/> ISLR	Data for an Introduction to Statistical Learning with Applications in R	1.2
System Library		
<input checked="" type="checkbox"/> base	The R Base Package	3.6.2
<input type="checkbox"/> boot	Bootstrap Functions (Originally by Angelo Canty for S)	1.3-23
<input type="checkbox"/> class	Functions for Classification	7.3-15
<input type="checkbox"/> cluster	"Finding Groups in Data": Cluster Analysis Extended Rousseeuw et al.	2.1.0
<input type="checkbox"/> codetools	Code Analysis Tools for R	0.2-16
<input type="checkbox"/> compiler	The R Compiler Package	3.6.2
<input type="checkbox"/> datasets	The R Datasets Package	3.6.2
<input type="checkbox"/> foreign	Read Data Stored by 'Minitab', 'S', 'SAS', 'SPSS', 'Stata', 'Systat', 'Weka', 'dBase', ...	0.8-72
<input checked="" type="checkbox"/> graphics	The R Graphics Package	3.6.2
<input checked="" type="checkbox"/> grDevices	The R Graphics Devices and Support for Colours and Fonts	3.6.2
<input type="checkbox"/> grid	The Grid Graphics Package	3.6.2
<input type="checkbox"/> KernSmooth	Functions for Kernel Smoothing Supporting Wand & Jones (1995)	2.23-16
<input type="checkbox"/> lattice	Trellis Graphics for R	0.20-38
<input type="checkbox"/> MASS	Support Functions and Datasets for Venables and Ripley's MASS	7.3-51.4
<input type="checkbox"/> Matrix	Sparse and Dense Matrix Classes and Methods	1.2-18
<input checked="" type="checkbox"/> methods	Formal Methods and Classes	3.6.2
<input type="checkbox"/> mgcv	Mixed GAM Computation Vehicle with Automatic Smoothness Estimation	1.8-31
<input type="checkbox"/> nlme	Linear and Nonlinear Mixed Effects Models	3.1-142
<input type="checkbox"/> nnet	Feed-Forward Neural Networks and Multinomial Log-Linear Models	7.3-12
<input type="checkbox"/> parallel	Support for Parallel computation in R	3.6.2

Graphics

The screenshot displays the RStudio interface. The main editor window contains R code for generating random data and plotting it. The console at the bottom shows the execution of this code. On the right, the 'Environment' pane shows the current workspace, and the 'Plots' pane shows the documentation for the `seq` function.

```
1 ## Graphics
2
3 x <- rnorm(50)
4 y <- x + rnorm(50, mean = 50, sd = .1)
5
6 plot(x, y)
7 plot(x, y, xlab = "this is the x-axis",
8      ylab = "this is the y-axis",
9      main = "Plot of X vs Y")
10 pdf("Figure.pdf")
11 plot(x, y, col = "green")
12 dev.off()
13
14 ?seq
15 x <- 1:10
16 x
17 x <- seq(1, 10)
18 x
```

Console output:

```
> x <- rnorm(50)
> y <- x + rnorm(50, mean = 50, sd = .1)
>
> plot(x, y)
> plot(x, y, xlab = "this is the x-axis",
+      ylab = "this is the y-axis",
+      main = "Plot of X vs Y")
> pdf("Figure.pdf")
> plot(x, y, col = "green")
> dev.off()
RStudioGD
2
>
> ?seq
> x <- 1:10
> x
[1] 1 2 3 4 5 6 7 8 9 10
> x <- seq(1, 10)
> x
[1] 1 2 3 4 5 6 7 8 9 10
> |
```

Environment pane:

Values	
x	int [1:10] 1 2 3 4 5 6 7 8 9 10
y	num [1:50] 50.5 49.9 51.1 50.7 51.1 ...

Plots pane: R Documentation

Sequence Generation

Description

Generate regular sequences. `seq` is a standard generic with a default method. `seq.int` is a primitive which can be much faster but has a few restrictions. `seq_along` and `seq_len` are very fast primitives for two common cases.

Usage

```
seq(...)
```

Default S3 method:

```
seq(from = 1, to = 1, by = ((to - from)/(length.out - 1)),
    length.out = NULL, along.with = NULL, ...)
```

```
seq.int(from, to, by, length.out, along.with, ...)
```

```
seq_along(along.with)
seq_len(length.out)
```

Arguments

...	arguments passed to or from methods.
from, to	the starting and (maximal) end values of the sequence. Of length 1 unless just <code>from</code> is supplied as an unnamed argument.
by	number: increment of the sequence.
length.out	desired length of the sequence. A non-negative number, which for <code>seq</code> and <code>seq.int</code> will be rounded up if fractional.

Indexing Data

The screenshot displays the RStudio environment with three main panes: Source, Environment, and Console.

Source Pane: Contains R code for creating a matrix and performing various indexing operations.

```
1 # Indexing Data
2
3 A <- matrix(1:16,4,4)
4 A
5 A[2,3]
6 A[c(1,3),c(2,4)]
7 A[1:3,2:4]
8 A[1:2,]
9 A[,1:2]
10 A[1,]
11 A[-c(1,3),]
12 A[-c(1,3),-c(1,3,4)]
13 dim(A)
14
15
16
```

Environment Pane: Shows the Global Environment with variables A, x, and y.

Variable	Class	Values
A	int [1:4, 1:4]	1 2 3 4 5 6 7 8 9 10 ...
x	int [1:10]	1 2 3 4 5 6 7 8 9 10
y	num [1:50]	50.5 49.9 51.1 50.7 51.1 ...

Console Pane: Shows the output of the R code execution.

```
[1] 1 2 3 4 5 6 7 8 9 10
> # Indexing Data
>
> A <- matrix(1:16,4,4)
> A
      [,1] [,2] [,3] [,4]
[1,] 1    5    9   13
[2,] 2    6   10   14
[3,] 3    7   11   15
[4,] 4    8   12   16
> A[2,3]
[1] 10
> A[c(1,3),c(2,4)]
      [,1] [,2]
[1,] 5    13
[2,] 7    15
> A[1:3,2:4]
      [,1] [,2] [,3]
[1,] 5    9   13
[2,] 6   10   14
[3,] 7   11   15
> A[1:2,]
      [,1] [,2] [,3] [,4]
[1,] 1    5    9   13
[2,] 2    6   10   14
```

Additional Graphical and Numerical Summaries

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

```
9 contour(x,y,T,nlevels=45,add=T)
10 fa <- (f-t(f))/2
11 contour(x,y,fa,nlevels=15)
12 image(x,y,fa)
13 persp(x,y,fa)
14 persp(x,y,fa,theta=30)
15 persp(x,y,fa,theta=30,phi=20)
16 persp(x,y,fa,theta=30,phi=70)
17 persp(x,y,fa,theta=30,phi=40)
18
19
20 plot(cylinders, mpg)
21 plot(Auto$cylinders, Auto$mpg)
22 attach(Auto)
23 plot(cylinders, mpg)
24 cylinders <- as.factor(cylinders)
25 plot(cylinders, mpg)
26 plot(cylinders, mpg, col="red")
27 plot(cylinders, mpg, col="red", varwidth=T)
28 plot(cylinders, mpg, col="red", varwidth=T, horizontal=T)
29 plot(cylinders, mpg, col="red", varwidth=T, xlab="cylinders", ylab="MPG")
30 hist(mpg)
31 hist(mpg,col=2)
32 hist(mpg,col=2,breaks=15)
33 pairs(Auto)
34 pairs(~ mpg + displacement + horsepower + weight + acceleration, Auto)
35 plot(horsepower,mpg)
36 summary(Auto)
37 summary(mpg)
38
39
40
```

Environment History Connections

Global Environment

Data

A	int [1:4, 1:4]	1 2 3 4 5 6 7 8 9 10 ...
f	num [1:50, 1:50]	-0.092 -0.0992 -0.1072 -0.1163 -0.1264 ...
fa	num [1:50, 1:50]	0 -0.00398 -0.00913 -0.0155 -0.02313 ...

Values

cylinders	Factor w/ 5 levels "3","4","5","6",...: 5 5 5 5 5 5 5 5 5 ...	
x	num [1:50]	-3.14 -3.01 -2.89 -2.76 -2.63 ...
y	num [1:50]	-3.14 -3.01 -2.89 -2.76 -2.63 ...

Files Plots Packages Help Viewer

Zoom Export Publish

Console

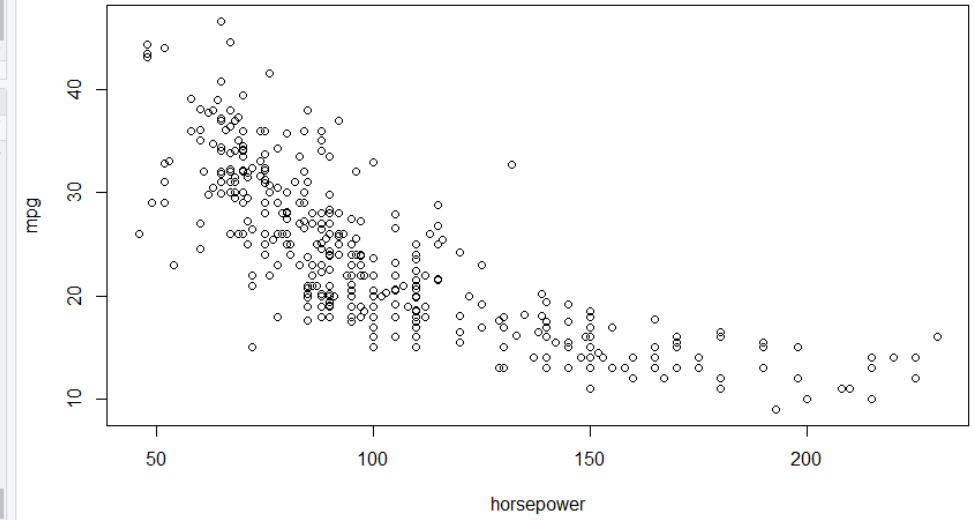
```
> plot(horsepower,mpg)
> summary(Auto)
```

	mpg	cylinders	displacement	horsepower	weight	acceleration	year
Min.	: 9.00	Min. :3.000	Min. : 68.0	Min. : 46.0	Min. :1613	Min. : 8.00	Min. :70.00
1st Qu.	:17.00	1st Qu.:4.000	1st Qu.:105.0	1st Qu.: 75.0	1st Qu.:2225	1st Qu.:13.78	1st Qu.:73.00
Median	:22.75	Median :4.000	Median :151.0	Median : 93.5	Median :2804	Median :15.50	Median :76.00
Mean	:23.45	Mean :5.472	Mean :194.4	Mean :104.5	Mean :2978	Mean :15.54	Mean :75.98
3rd Qu.	:29.00	3rd Qu.:8.000	3rd Qu.:275.8	3rd Qu.:126.0	3rd Qu.:3615	3rd Qu.:17.02	3rd Qu.:79.00
Max.	:46.60	Max. :8.000	Max. :455.0	Max. :230.0	Max. :5140	Max. :24.80	Max. :82.00

	origin	name
Min.	:1.000	amc matador : 5
1st Qu.	:1.000	ford pinto : 5
Median	:1.000	toyota corolla : 5
Mean	:1.577	amc gremlin : 4
3rd Qu.	:2.000	amc hornet : 4
Max.	:3.000	chevrolet chevette: 4
		(other) :365


```
> summary(mpg)
```

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
mpg	9.00	17.00	22.75	23.45	29.00	46.60



A scatter plot showing the relationship between horsepower (x-axis) and miles per gallon (mpg, y-axis). The x-axis ranges from 50 to 250, and the y-axis ranges from 10 to 40. The plot shows a negative correlation, with many data points clustered between 50 and 150 horsepower and 15 to 35 mpg. There are a few outliers with high horsepower (above 200) and low mpg (below 15).