## Common questions (and answers) about R

David M. Rosenberg

Committee on Neurobiology University of Chicago

October 20, 2009



### Outline

- Admin issues
  - Things to ask for
  - Things to provide
- Manipulating Data Vectors
  - Basic usage
  - Indexing
- Complex types
  - Overview
  - Slicing and extracting
  - Syntax

- Functions
  - Unexplained functions
  - Reading the documentation
- Style
  - Submissions
  - Rules
  - The golden rule

#### Email addresses:

If you use a non-University of Chicago email account, let me know either via an email from your uchicago.edu email account or a written note.

#### Email addresses:

If you use a non-University of Chicago email account, let me know either via an email from your uchicago.edu email account or a written note.

### Use of lab time:

- Status quo. Labs are principally time for you to try to work on the lab.
- Short lectures. Presentations like this (but shorter; 10-20 minutes) at the start / end / middle of labs seems helpful.
- Review of the previous week's assignment and / or group discussion of the different ways you solved the problems.
- Discussion / question-and-answer time about the topics and concepts presented in lecture not directly pertaining to computational programming.

#### Email addresses:

If you use a non-University of Chicago email account, let me know either via an email from your uchicago.edu email account or a written note.

#### Use of lab time:

- Status quo. Labs are principally time for you to try to work on the lab.
- Short lectures. Presentations like this (but shorter; 10-20 minutes) at the start / end / middle of labs seems helpful.
- Review of the previous week's assignment and / or group discussion of the different ways you solved the problems.
- Discussion / question-and-answer time about the topics and concepts presented in lecture not directly pertaining to computational programming



#### Email addresses:

If you use a non-University of Chicago email account, let me know either via an email from your uchicago.edu email account or a written note.

#### Use of lab time:

- Status quo. Labs are principally time for you to try to work on the lab.
- Short lectures. Presentations like this (but shorter; 10-20 minutes) at the start / end / middle of labs seems helpful.
- Review of the previous week's assignment and / or group discussion of the different ways you solved the problems.
- Discussion / question-and-answer time about the topics and concepts presented in lecture not directly pertaining to computational programming



#### Email addresses:

If you use a non-University of Chicago email account, let me know either via an email from your uchicago.edu email account or a written note.

#### Use of lab time:

- Status quo. Labs are principally time for you to try to work on the lab.
- Short lectures. Presentations like this (but shorter; 10-20 minutes) at the start / end / middle of labs seems helpful.
- Review of the previous week's assignment and / or group discussion of the different ways you solved the problems.
- Discussion / question-and-answer time about the topics and concepts presented in lecture not directly pertaining to computational programming



#### Email addresses:

If you use a non-University of Chicago email account, let me know either via an email from your uchicago.edu email account or a written note.

#### Use of lab time:

- Status quo. Labs are principally time for you to try to work on the lab.
- Short lectures. Presentations like this (but shorter; 10-20 minutes) at the start / end / middle of labs seems helpful.
- Review of the previous week's assignment and / or group discussion of the different ways you solved the problems.
- Discussion / question-and-answer time about the topics and concepts presented in lecture not directly pertaining to computational programming.

- Homework examples & template:
   Files on chalk.
- Grades & comments:
   By next week you will have exercises 0 and 1 returned to you.
- Tools & utilities:
   Source code "cleaner," problem uploader
- Cheat sheets:
   Math "cheat sheets" (i.e. trigonometric identities, tables of common integrals and derivatives, etc.) as well as "cheat sheets" for R are in progress.

- Homework examples & template:
   Files on chalk.
- Grades & comments:
  - By next week you will have exercises 0 and 1 returned to you.
- Tools & utilities:
   Source code "cleaner," problem uploader
- Cheat sheets:

Math "cheat sheets" (i.e. trigonometric identities, tables of common integrals and derivatives, etc.) as well as "cheat sheets" for *R* are in progress.

- Homework examples & template:
   Files on chalk.
- Grades & comments:
   By next week you will have exercises 0 and 1 returned to you.
- Tools & utilities:
   Source code "cleaner," problem uploader
- Cheat sheets:
   Math "cheat sheets" (i.e. trigonometric identities, tables of common integrals and derivatives, etc.) as well as "cheat sheets" for R are in progress.

- Homework examples & template:
   Files on chalk.
- Grades & comments:
   By next week you will have exercises 0 and 1 returned to you.
- Tools & utilities: Source code "cleaner," problem uploader
- Cheat sheets:
  - Math "cheat sheets" (i.e. trigonometric identities, tables of common integrals and derivatives, etc.) as well as "cheat sheets" for  $\it R$  are in progress.

### Outline

- Admin issues
  - Things to ask for
  - Things to provide
- Manipulating Data Vectors
  - Basic usage
  - Indexing
- Complex types
  - Overview
  - Slicing and extracting
  - Syntax

- Functions
  - Unexplained functions
  - Reading the documentation
- Style
  - Submissions
  - Rules
  - The golden rule

# Manipulating Data - Vectors Basic usage

#### Rules

- Single type
- Accessed by index
- Can be named or unnamed

```
>c('hello', TRUE); c(1, TRUE)
[1] "hello" "TRUE"
[1] 11
>class(c('hello', TRUE)); class(c(1, TRUE));
[1] "character"
>c('Hello', 14);
[1] "Hello" "14"
>class(c(TRUE, TRUE));
[11 "logical"
```

```
>my_name <- c('D', 'a', 'v', 'i', 'd', '', 'R', '.');
>length (my name);
[1] 8
>names(my name);
NULL
>names(mv name) <-
   c( paste('first', as.character(1:5),
           sep="_"),
       'space'.
       paste ('last', as.character (1:2),
             sep=" ")
    );
>my name;
first 1 first 2 first 3 first 4 first 5
            "a" "v" "i"
  space last 1 last 2
```

## Manipulating Data - Vectors Basic usage

### Altering vectors

## Use <- to alter / overwrite an existing vector

### Growing vectors

```
Use c (vector, othervector, value, ...) to add to an existing vector
```

### Removing elements

```
Use vec <- vec[-(element_index)] to remove the element at element_index from vec
```

### Single-element alterations

```
Use vec[index] <- newvalue to alter a
single element</pre>
```

Beware of type mismatches here!

```
>temp vec <- 1:5;
>temp vec
[11 1 2 3 4 5
>temp vec <- c(temp vec. 6):
>temp_vec:
[1] 1 2 3 4 5 6
>temp vec2 <- c(temp vec, 'Cat');
>temp_vec2;
[1] "1" "2"
[7] "Cat"
>temp_vec[-2];
[1] 1 3 4 5 6
>temp vec[3] <- 100;
>temp vec;
[1] 1 2 100 4
>temp_vec1 <- 1:5: temp_vec2 <- 7:10:
>c(temp_vec1, 6, temp_vec2);
[1] 1 2 3 4 5 6 7 8 9 10
```

## Manipulating Data - Vectors

### Numerical / nominative

- Ranges
- By names
- Finding indexes

```
>my name[1:3];
first 1 first 2 first 3
    "D" "a" "v"
>mv name[(1:3) * 21:
first 2 first 4 space
   "a" "i" " "
>my name[c(1, 5, length(my name))];
first 1 first 5 last 2
   "D" "d"
>my_name['first_3'];
first 3
   "v"
>my name[c('first 1', 'last 1')]
first 1 last 1
   "D"
           "R"
```

### Logical

Review

```
>temp vec <- 1:10;
>temp vec <= 5;
[1] TRUE TRUE TRUE TRUE TRUE FALSE
[7] FALSE FALSE FALSE FALSE
>temp vec[temp vec <= 5]:
[1] 1 2 3 4 5
>my name == 'D'
first_1 first_2 first_3 first_4 first_5
  TRUE FALSE FALSE FALSE
 space last 1 last 2
 FAISE FAISE FAISE
>my_name[my_name == 'D'];
first 1
    "D"
>my name[my name == ' ' | my name == '.'];
space last 2
>my_name[!(my_name %in% c(' ', '.'))];
first 1 first 2 first 3 first 4 first 5
   "D" "a" "v" "i"
last 1
    "R"
```

### Outline

- Admin issues
  - Things to ask for
  - Things to provide
- Manipulating Data Vectors
  - Basic usage
  - Indexing
- Complex types
  - Overview
  - Slicing and extracting
  - Syntax

- Functions
  - Unexplained functions
  - Reading the documentation
- Style
  - Submissions
  - Rules
  - The golden rule

## Complex types Overview

#### Lists

Lists are like vectors, except that they can hold *different* types in different slots.

```
>list_ex2;

$name_ta

first_1 first_2 first_3 first_4 first_5

"D" "a" "v" ";" "d"

space last_1 last_2

" "R" "

$one_to_five

[1] 1 2 3 4 5

$mytun

function(x) { return(1/x) }

>class(list_ex);

[1] "list"

>class(list_ex2[[1]]);

[1] "character"
```

## Complex types

#### Data frames

Data frames are kind of like matrices and kind of like lists. Different members can be of different types (like lists). Each column (member) of a data frame, however, must be of equal length (like columns in a matrix).<sup>a</sup>

<sup>a</sup>In fact, data frames are lists.

```
>cost per=c(0.10, 0.25, 0.50);
>items=c('spam', 'eqq', 'foobar');
>on hand=c(123, 153, 55);
>df <- data.frame(items=items, on hand=on hand.
       cost per=cost per):
>df
   items on hand cost per
    spam
             123
                      0.10
                      0.25
     eaa
             153
3 foobar
              55
                      0.50
>total value <- df$on hand * df$cost per
```

```
>df <- cbind(df, total value=total value)
>df
   items on hand cost per total value
    spam
             123
                      0.10
                                  12 30
     eaa
             153
                      0.25
                                  38.25
3 foobar
              55
                      0.50
                                  27.50
>df2 <- cbind(df, weight per=c(0.5, 0.1, 3))
>df2
   items on hand cost per total value
              123
                      0.10
                                  12.30
    spam
                      0.25
     egg
              153
                                  38.25
3 foobar
              55
                      0.50
                                  27.50
  weight per
         0.5
         0.1
         3.0
```

## Complex types Overview

```
>df2 <- cbind(
       df2.
       total weight=df2$on hand * df2$weight per.
       value density=df2$total value/(df2$on hand * df2$weight per)
>df2
   items on_hand cost_per total_value
    spam
             123
                     0.10
                                 12.30
             153
                     0.25
2
     egg
                                 38.25
3 foobar
              55
                     0.50
                                 27.50
  weight per total weight value density
         0.5
                     61.5
                              0.2000000
2
         0.1
                     15.3
                               2.5000000
         3.0
                    165.0
                               0.1666667
>class(df2);
[1] "data.frame"
>is.list(df2);
[1] TRUE
```

## Complex types Slicing and extracting

### Slicing

Using brackets ([indices]-notation) to create "subsets" of a vector, list, or data frame is called *slicing*.

### Slice type preservation

Slicing an object always returns an object of the same type.

- A slice of a character vector is a character vector; a slice of a numeric vector is a numeric vector, etc.
- A slice of a list / data frame is always a list / data frame (even if it has only 1 member).
- A SLICE OF A LIST / DATA FRAME IS ALWAYS A LIST / DATA FRAME (EVEN IF IT HAS ONLY 1 MEMBER).



## Complex types Slicing and extracting

## Extracting

Using double-brackets ([[index]]) or dollar-sign notation (mylistSmember\_name) to examine single members of a list or data frame is called extraction.

### Extract types

The result of an extraction operation is of type defined by the returned member.

- Extraction only works with lists / data frames.
- Extraction can only return a single member.



#### Slice / extraction

- Slicing can only be performed using single brackets.
- Extraction can only be performed using double-brackets or dollar-sign - name notation.

- cbind() joins matrices / data frames columnwise
- rbind() joins matrices / data frames columnwise
- names() is used for getting and setting element names
- attr() is used for getting and setting other object attributes.

```
>df <- data.frame(items=items, on hand=on hand,
       cost per=cost per);
>list1 <- as.list(df);
>list1:
$items
[1] spam
           eaa
                  foobar
Levels: egg foobar spam
$on hand
[11 123 153 55
$cost per
[1] 0.10 0.25 0.50
>names(df);
[1] "items"
               "on hand" "cost per"
>names(df)[1] <- c('Mv items'):
 My items on_hand cost_per
      spam
               123
                        0.10
                153
                        0.25
       eaa
   foobar
                55
                        0.50
```

#### Slice / extraction

- Slicing can only be performed using single brackets.
- Extraction can only be performed using double-brackets or dollar-sign - name notation.

- cbind() joins matrices / data frames columnwise
- rbind() joins matrices / data frames columnwise
- names () is used for getting and setting element names
- attr() is used for getting and setting other object attributes.

```
>df <- data.frame(items=items, on hand=on hand,
       cost per=cost per);
>list1 <- as.list(df);
>list1:
$items
[1] spam
           eaa
                  foobar
Levels: egg foobar spam
$on hand
[11 123 153 55
$cost per
[1] 0.10 0.25 0.50
>names(df);
[1] "items"
               "on hand" "cost per"
>names(df)[1] <- c('Mv items'):
>df
 My items on_hand cost_per
      spam
               123
                        0.10
                153
                        0.25
       eaa
   foobar
                55
                        0.50
```

#### Slice / extraction

- Slicing can only be performed using single brackets.
- Extraction can only be performed using double-brackets or dollar-sign - name notation.

- cbind() joins matrices / data frames columnwise
- rbind() joins matrices / data frames columnwise
- names() is used for getting and setting element names
- attr() is used for getting and setting other object attributes.

```
>df <- data.frame(items=items, on hand=on hand,
       cost per=cost per);
>list1 <- as.list(df);
>list1:
$items
[1] spam
           eaa
                  foobar
Levels: egg foobar spam
$on hand
[1] 123 153 55
$cost per
[1] 0.10 0.25 0.50
>names(df);
[1] "items"
               "on hand" "cost per"
>names(df)[1] <- c('Mv items'):
 My items on hand cost per
      spam
               123
                        0.10
                        0.25
       eaa
   foobar
                55
                        0.50
```

### Slice / extraction

- Slicing can only be performed using single brackets.
- Extraction can only be performed using double-brackets or dollar-sign - name notation.

- cbind() joins matrices / data frames columnwise
- rbind() joins matrices / data frames columnwise
- names() is used for getting and setting element names
- attr() is used for getting and setting other object attributes.

```
>df <- data.frame(items=items, on hand=on hand,
       cost per=cost per);
>list1 <- as.list(df);
>list1:
$items
[1] spam
           eaa
                  foobar
Levels: egg foobar spam
$on hand
[1] 123 153 55
$cost per
[1] 0.10 0.25 0.50
>names(df);
[1] "items"
               "on hand" "cost per"
>names(df)[1] <- c('Mv items'):
 My items on hand cost per
      spam
               123
                        0.10
                153
                        0.25
       eaa
   foobar
                55
                        0.50
```

#### Slice / extraction

- Slicing can only be performed using single brackets.
- Extraction can only be performed using double-brackets or dollar-sign - name notation.

- cbind() joins matrices / data frames columnwise
- rbind() joins matrices / data frames columnwise
- names() is used for getting and setting element names
- attr() is used for getting and setting other object attributes.

```
>df <- data.frame(items=items, on hand=on hand,
       cost per=cost per);
>list1 <- as.list(df);
>list1:
$items
[1] spam
           eaa
                  foobar
Levels: egg foobar spam
$on hand
[1] 123 153 55
$cost per
[1] 0.10 0.25 0.50
>names(df);
[1] "items"
               "on hand" "cost per"
>names(df)[1] <- c('Mv items'):
 My items on hand cost per
      spam
               123
                        0.10
                153
                        0.25
       eaa
   foobar
                55
                        0.50
```

#### Slice / extraction

- Slicing can only be performed using single brackets.
- Extraction can only be performed using double-brackets or dollar-sign - name notation.

- cbind() joins matrices / data frames columnwise
- rbind() joins matrices / data frames columnwise
- names() is used for getting and setting element names
- attr() is used for getting and setting other object attributes.

```
>df <- data.frame(items=items, on hand=on hand,
       cost per=cost per);
>list1 <- as.list(df);
>list1:
$items
[1] spam
                  foobar
           eaa
Levels: egg foobar spam
$on hand
[11 123 153 55
$cost per
[1] 0.10 0.25 0.50
>names(df);
[1] "items"
               "on hand" "cost per"
>names(df)[1] <- c('Mv items'):
 My items on_hand cost_per
      spam
               123
                        0.10
                153
                        0.25
       eaa
   foobar
                55
                        0.50
```

### Outline

- Admin issues
  - Things to ask for
  - Things to provide
- Manipulating Data Vectors
  - Basic usage
  - Indexing
- Complex types
  - Overview
  - Slicing and extracting
  - Syntax

- Functions
  - Unexplained functions
  - Reading the documentation
- Style
  - Submissions
  - Rules
  - The golden rule

### Reading other people's functions

If you *need* to know how the functions I gave you work, here's an overview of what I've skipped over and why.

- random data The functions runif() and rnorm() are used to generate arbitrary random numeric vectors.
- ② **string manipulation** Manipulation of strings uses *another* language called *regular expressions* or *regex* that is embedded in *R* (and nearly all other languages). Everything i've given you that manipulates text (e.g. <a href="mailto:parsePolynomial">parsePolynomial</a>()) uses a number of *regex* functions including <a href="mailto:gsub()">gsub()</a>, <a href="mailto:strsplit">strsplit</a>() and <a href="mailto:passe()">passe()</a>. String processing is a great thing to learn, but not the focus of this course.

### Reading other people's functions

If you *need* to know how the functions I gave you work, here's an overview of what I've skipped over and why.

- other Most of the remaining functions that you have seen but not learned about are either
  - Peculiarities of the R language and not useful in terms of general numerical computing. (e.g. options (expressions=500000);)
  - Operating system interaction. (e.g. system.time(qSort(inVec));)
  - Artifacts of my formatting. (e.g. the first "block" in the source I posted for appendix 1.)
  - Abstract / challenging / confusing in nature or implementation. (The hacks used in the maxima\_utilites.R file)



## Functions Reading the documentation

## Documentation

Consult the online documentation frequently. This includes the examples.

Invariably, you will encounter both functions that you have forgotten how to use and "problems" for which you would *expect* a function to have already addressed.

How do you expect them to graph the functions for the cobweb plots? Is there an R function for plotting a given algebraic expression? This also applies for plotting the exponential functions they generate in the glucose model question.

Look at the code on the top left of page 3. I think this should answer that question. There is another function, curve() that I haven't talked about that can be used as well. Try the following:

```
curve (1-\exp(-x), \text{ from}=0.01, \text{ to}=10)
```



## Functions Reading the documentation

Given the assumption that generation a plot from function might be in the "graphics" package . . .



## **Functions**

Reading the documentation

- help() (?) Function documentation
- help(package=package name)
- help.search() (??)
- help.start(
- example()
- function name (without
  parenthesis)

```
>help(ls);
>?ls
>?`+` # use backquotes when R doesn't
```

## Functions Reading the documentation

- help()(?)
- help (package=package name) - Package documentation listing
- documentation listing
- help.search() (??)
- ncip.scarc
- function name (Withou

- Package: Rinterval Type: Package
- Title: Package for efficiently dealing with
- genomic intervals
  Version: 0.14
- Date: 2009-09-09
- Author: David M. Rosenberg Maintainer: Who to complain to
- <rosenbergdm@uchicago.edu>
- Description: Classes and methods for efficiently analyzing single dimension genomic
  - intervals, with an emphasis on copy
  - number variation.
- License: LGPL LazyLoad: yes
- Packaged: 2009-09-10 17:19:11 UTC; root
- Built: R 2.9.2; ; 2009—10—15 17:33:27 UTC;
  - unix



## **Functions**

Reading the documentation

### **Commands**

```
help() (?)help(package=package
```

 help.search() (??) -Search all documentation for occurances of the given term.

```
help.start()
example()
function_name(W
```

function name (without parenthesis)

```
>help.search('digest');
```

## **Functions**

Reading the documentation

- help()(?)
- help(package=package
  name)
- help.search() (??)
- help.start() Open HTML help browser in the default web browser.
- example()
- function name (without parenthesis)

## Functions Reading the documentation

- help() (?)
- help(package=package name)
- help.search() (??)
- help.start()
- example() Execute any example code in a function's documentation. These examples should never fail.
- function name (without parenthesis)

```
>example(is.integer);
is.ntp> ## as.integer() truncates:
is.ntp> x <- pi * c(-1:1,10)
is.ntp> as.integer(x)
[1] -3 0 3 31
```

## **Functions**

Reading the documentation

- help()(?)
- help(package=package name)
- help.search() (??)
- help.start()
- example()
- function name (without parenthesis) - Show function body.

```
sisTRUE;
function (x)
identical (TRUE, x)
<environment: namespace:base>
>lsf.str;
function (pos = -1, envir, ...)
{
   if (missing(envir))
       envir <- as.environment(pos)
   ls.str(pos = pos, envir = envir, mode = "function", ...)
}
<environment: namespace:utils >
```

### Outline

- Admin issues
  - Things to ask for
  - Things to provide
- Manipulating Data Vectors
  - Basic usage
  - Indexing
- Complex types
  - Overview
  - Slicing and extracting
  - Syntax

- Functions
  - Unexplained functions
  - Reading the documentation
- Style
  - Submissions
  - Rules
  - The golden rule

### Example and template

An example and a template for how I'd like you to *try* to submit your homework to me will be available by tonight.

The "submission" style guidelines are primarily to speed my grading of homework

- Each assignment should be submitted as a single .R file. Additional (non-.R) files / hard copies may be necessary, including
  - Plots
  - Analytical problems
  - Data tables
- Anything in the assignment which is not R code should be commented out (preceded by pound signs).



#### File header

The following header should start your . R file (the first three lines may be omitted). Substitute the relevant information for each all-caps slot.

```
#!/usr/bin/env r
# encoding: utf-8
#
name: STUDENT_NAME
# assignment: ASSIGNMENT_NUMBER
# date: DATE_OF_SUBMISSION
# filename: NAME OF FILE
```

#### Question heade

The following block should be inserted before each answer

```
# name: STUDENT_NAME
# assignment: ASSIGNMENT_NUMBER
# date: DATE_OF_SUBMISSION
# question: QUESTION_NUMBER
# subquestion: SUBQUESTION_LETTER_IF_PRESENT_OTHERWISE_EMPTY
# other files: NAMES_OF_OTHER_RELATED_FILES
```

#### File header

The following header should start your .R file (the first three lines may be omitted). Substitute the relevant information for each all-caps slot.

### Question header

The following block should be inserted before each answer.

```
# name: STUDENT_NAME
# assignment: ASSIGNMENT_NUMBER
# date: DATE_OF_SUBMISSION
# question: QUESTION_NUMBER
# subquestion: SUBQUESTION_LETTER_IF_PRESENT_OTHERWISE_EMPTY
# other files: NAMES OF OTHER RELATED FILES
```

Category	Description
variable names	Variable names should describe their function and follow
indentation	a consistent capitalization scheme code blocks (generally delimited by brackets) should be indented (recommended 2 spaces) relative to surrounding code
blank lines	use blank lines to separate independent "chunks" or concepts
spaces	blank spaces should surround symbols and parentheses
comments	comments should be used liberally to explain code and concepts
line length	Individual lines should be no more than 80 characters long. Continuation lines (if necessary) should be indented relative to surrounding code

### Style The golden rule

## The Golden Rule

Whatever your personal stylistic preferences / conventions are, **be consistent**.