UNIVERSITY OF CAMBRIDGE

Bioinformatics Training

Introduction to R

for Biologists

9 -10 September 2019

Bioinformatics Training Room, Downing Site, Cambridge

Instructors: Karsten Bach, Alexia Cardona, Hugo Tavares

Data types (recap)

- **logical**: TRUE FALSE
- **integer**: whole numbers. e.g., 40
- **double**: numbers with decimal points. e.g., 2.666
- **character**: words or **strings**. e.g., "Hello"

Logical operators

&	and
|	or
!	not
==	equal
>	greater than
<	less than
>=	greater than or equal to
<=	less than or equal to
!=	not equal to

Data structures (recap)

- **vector**: list of items of the same data type. E.g., c(4, 6, 9,12)
- **factor**: categorical data (has to be a character vector).
- **data.frame**: contains tabular data – normally data is loaded into data.frame when reading in a file.

Data structures (recap)

- **vector**: list of items of the same data type. E.g., c(4, 6, 9,12)

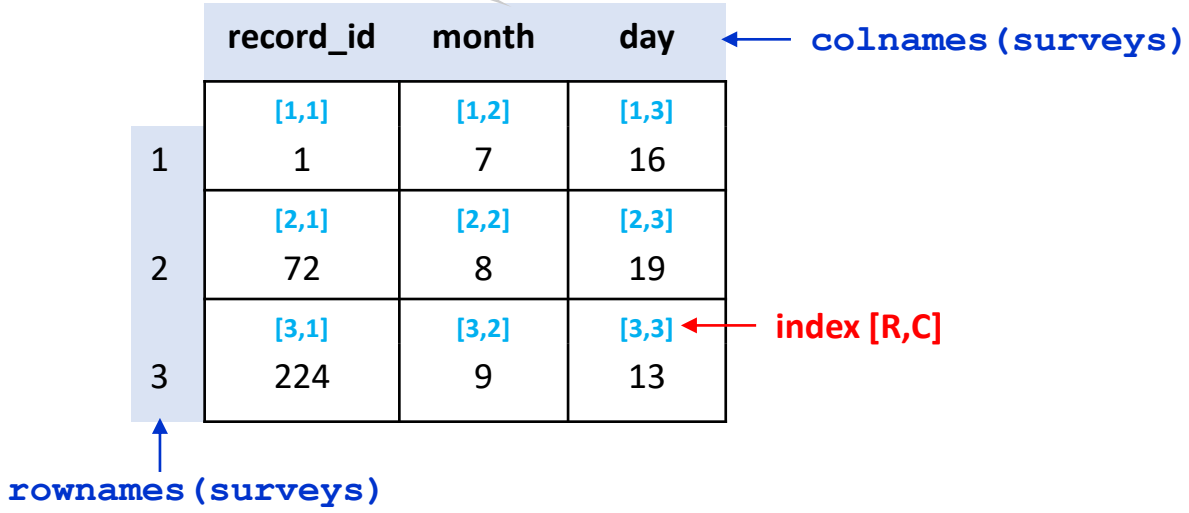
- **factor**: categorical data (has to be a character vector).

  See https://cambiotraining.github.io/IntroR_9Sep2019/02-starting-with-data.html#factors

- **data.frame**: contains tabular data – normally data is loaded into data.frame when reading in a file.

Data structures (recap)

```
> head(surveys)
  record_id month day year plot_id species_id sex hindfoot_length weight   genus  species   taxa plot_type
1         1     7  16 1977       2         NL   M              32     NA Neotoma albigula Rodent   Control
2        72     8  19 1977       2         NL   M              31     NA Neotoma albigula Rodent   Control
3       224     9  13 1977       2         NL                  NA     NA Neotoma albigula Rodent   Control
4                  [illegible]    2         NL                  NA     NA Neotoma albigula Rodent   Control
5                  [illegible]    2         NL                  NA     NA Neotoma albigula Rodent   Control
6       363        [illegible]    2         NL                  NA     NA Neotoma albigula Rodent   Control
>
```



Numeric indexing:
surveys[1,2]

Name indexing (for columns):
surveys$month

Logical indexing:
surveys[surveys$month == 6]

Variables (recap)

- Containers of data types and data structures.

e.g.,

```
weight <- 55   #here weight is an integer
weight <- surverys$weight  # here weight is a vector
```