Data Types and Vectors

Intro to Programming in R
Week 2

Myfanwy Johnston, PhD

Data Types and Vectors

Source materials:

Essential R by Matthew B. Espe, PhD
Intro to R course by Anna Steel, PhD, UC Davis 2013

1. Data Types in R

Why do we need to care about data types in R?

Because our computers care.

Data Types in R

numeric/double	1.0, 507.345, 5e3, 67
integer	1, 2, 3, 5e3, 67
character	“dog”, “five”, “5”
logical	TRUE, FALSE

Every time we ask R to store a piece of data on our computer, R has to pre-allocate the correct amount of memory for that data.

Because different data types require different amounts of memory, R needs to coerce data of mixed type into a single type before storing it.

This system of pre-allocation, coercion, and memory storage is what makes R’s programmatic manipulations of data possible.

This coercion system can be confusing at first - TRUEs turn into 1s, or a bunch of integers turn into a bunch of characters, seemingly without our direction, permission, or (danger zone) our notice.

2. Vectors

Vector [**vek**-tər] (*noun*): a collection of data of uniform type; the simplest (lowest-level) data structure in R, from which many other structures are created.

```
my_num_vector = c(1.0, 5e5, 2.456, 67.0)
my_char_vector = c(“cat”, “dog”, “fish”)
my_data = data.frame(my_num_vector, my_char_vector)
```

When you try to mix different data types together in the same vector, R will coerce the whole vector to the same atomic type, according to a hierarchy*.

```
(*Most of the time, stuff
gets coerced to character)
```

Lesson Objectives:

1. Be able to define the word “vector” as it relates to R
2. Learn how to use the c() function to build separate vectors (one of each atomic data type) and save them to named objects in R.
3. Learn how to use the `[` function to extract different elements of each vector
4. Discover the data type(s) to which different mixed vectors are coerced when you try to put them together in the same vector.

Cheatsheet: creating four different atomic vectors

Type	R Code
numeric/double	`c(1.0, 507.345, 5e3, 67)`
integer	`c(1L, 2L, 3L, 5e3L, 67L)`
character	`c("cat", "dog")`
logical	`c(TRUE, FALSE)`

Course GitHub repository:
https://github.com/fishsciences/2020-R-Course