Vectors and Factors

MSDS 597 Data Wrangling & Husbandry

February 03, 2020

Vectors

Vectors in R are what a computer scientist might call 1-dimensional arrays. That is, a set of numbers, or a set of character strings, or a set of TRUE/FALSE Booleans, and so forth. They all have to have the same data type.

The simplest way to make them is using the `c()` function:

```
ex1 <- c(1, 7, 20, 2)
ex2 <- c("Rutgers", "Michigan", "Northwestern")
ex3 <- c(TRUE, TRUE)
ex4 <- c(1.234, 4.923, 7.3422, 8, -9.4)
```

R will change the data type of all of the elements if need be.

```
c(1,3,6, "ten", FALSE)
```

```
## [1] "1"     "3"     "6"     "ten"   "FALSE"
```

You can pick out an element using []:

```
ex4[3]
```

```
## [1] 7.3422
```

```
ex2[c(1,3)]
```

```
## [1] "Rutgers"      "Northwestern"
```

```
ex4[ex4 > 5]
```

```
## [1] 7.3422 8.0000
```

There are some very useful functions that generate vectors

```
3:7
```

```
## [1] 3 4 5 6 7
```

```
3:(-2)
```

```
## [1]  3  2  1  0 -1 -2
```

```
seq(1, 3, by = .2)
```

```
##  [1] 1.0 1.2 1.4 1.6 1.8 2.0 2.2 2.4 2.6 2.8 3.0
```

```r
rep("Rutgers", 5)
```

```
## [1] "Rutgers" "Rutgers" "Rutgers" "Rutgers" "Rutgers"
```

```r
rep(1:4, each = 2, times = 3)
```

```
##  [1] 1 1 2 2 3 3 4 4 1 1 2 2 3 3 4 4 1 1 2 2 3 3 4 4
```

Factors

- Factors cause me more trouble than any data type other than dates.
- A factor vector can be thought of as a categorical variable
  - A bit like like a vector of character strings
  - Structured differently
  - Essentially a vector of *integers* where each integer has a label

```
campus <- factor(c("Busch", "Busch", "Livingston", "College
                   "Douglass", "Cook", "Livingston", "Busch
campus
```

```
## [1] Busch       Busch       Livingston  College Ave Doug
## [7] Livingston  Busch
## Levels: Busch College Ave Cook Douglass Livingston
```

```
as.integer(campus)
```

```
## [1] 1 1 5 2 4 3 5 1
```

You can specify the order of levels if you like; the default is alphabetical

```
(campus.alt <- factor(c("Busch", "Busch", "Livingston", "Co
          "Douglass", "Cook", "Livingston", "Busch"),
       levels = c("Busch", "Livingston", "College Ave", "Do
```

```
## [1] Busch       Busch       Livingston  College Ave Doug
## [7] Livingston  Busch
## Levels: Busch Livingston College Ave Douglass Cook
```

```
as.integer(campus.alt)
```

```
## [1] 1 1 2 3 4 5 2 1
```

```
as.integer(campus)
```

```
## [1] 1 1 5 2 4 3 5 1
```

Further notes

- The default option for `read.table()` and `data.frame()` (in base R) and its variants is to convert characters to factors, which many people find annoying. You can turn that off by using the option `stringsAsFactors = FALSE`.
- The default option for `read_table()` and `tibble()`(in the tidyverse) and its variants is to keep characters as characters.
- The default for
- To convert a factor vector to a character vector, use

```
as.character(campus)
```

```
## [1] "Busch"       "Busch"       "Livingston"  "College A
## [6] "Cook"        "Livingston"  "Busch"
```