

# Few concepts in R

Anoop Mayampurath

January 9, 2019

## RStudio basics

- Download link: <https://www.rstudio.com/products/rstudio/download/>
- Installing Packages: Select Tools -> Install Packages. Type in the package you need and be sure to check the Install Dependencies button.
- You will need to install the following packages : plyr, dplyr, and lubridate before next class

## R Variables

The following are common variable types used in R

### Single-element variables

These variables only containing a single value

- integer, used to represent integer numbers - double, used to represent floating point numbers - character, used to represent single characters

On the console window, type in the following

```
x_integer <- 4 # example of an integer variable
x_double <- 4.5 # example of a double
x_char <- 'a' # example of a character variable
```

The above represent examples of single-element variables. You can also see their values on the Environment tab (top-right)

### Multi-element variables

These elements contain more than one variable.

- vectors
- factors
- lists

Vectors are one-dimensional variables containing elements of the same type (homogeneous)

```
x_int <- c(1,2,3) # x_int is an integer vector
x_int
```

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

```
x_double <- c(4.5, 7.8, 9, 10.6, 3.145) # example of a double vector
x_double
```

```
## [1] 4.500 7.800 9.000 10.600 3.145
```

```
x_char <- c("a", "b", "c", "d") # example of character vector
x_char
```

```
## [1] "a" "b" "c" "d"
```

Indexing vectors (i.e. accessing elements in a vector) is done using square brackets. Remember that R indexing starts at 1.

```
x_int[2]
```

```
## [1] 2
```

```
x_char[4]
```

```
## [1] "d"
```

```
x_double[1]
```

```
## [1] 4.5
```

You can use the function `length()` to get the number of elements in a vector

```
length(x_int)
```

```
## [1] 3
```

Factors are similar to vectors except they are categorical variables.

```
# Example of a factor with two levels
```

```
x_factor <- factor(c("Male", "Female", "Female", "Male"))
```

```
x_factor
```

```
## [1] Male   Female Female Male
```

```
## Levels: Female Male
```

```
levels(x_factor)
```

```
## [1] "Female" "Male"
```

You can convert other variables into factors using the `factor()` function.

```
x_char <- c("Male", "Female", "Male")
```

```
x_factor <- factor(x_char)
```

```
x_factor
```

```
## [1] Male   Female Male
```

```
## Levels: Female Male
```

```
x_int <- c(1, 2, 3, 1, 3)
```

```
x_factor <- factor(x_int)
```

```
x_factor
```

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

```
## Levels: 1 2 3
```

Lists are 1-D variables where elements can be of different types. They are considered heterogenous.

```
x_list <- list(c(1,2,3),  
              c("h", "a"),  
              c(1.4, 5.6, 7.8, 9.0),  
              "temp")
```

```
x_list
```

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

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

```
##
```

```
## [[2]]
## [1] "h" "a"
##
## [[3]]
## [1] 1.4 5.6 7.8 9.0
##
## [[4]]
## [1] "temp"
```

Indexing in lists are done using `[[ ]]` as well as `[ ]`.

```
x_list[[1]]
```

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

```
x_list[[4]]
```

```
## [1] "temp"
```

```
x_list[[3]][1] # for the 1st entry within the 3rd element of the list
```

```
## [1] 1.4
```

```
x_list[[1]][2]
```

```
## [1] 2
```

## Data Frames

Data frames are 2-D variables that combines vectors of different types as columns. In this class, we will use `read.csv` to load data into a data frame.

```
# Reading in a dataset into a data frame
```

```
d.in <- read.csv("~/Google Drive/teaching/2019_winter_msbi_32000/lectures/lecture_2/patient_dataset.csv")
```

You can see a snapshot of the dataset

```
# First few rows
```

```
head(d.in)
```

```
##   patient.ids      patient.names      dob gender
## 1           1      Matthew Connor 8/12/1965    M
## 2           2        Joanna Lujan 7/26/1989    F
## 3           3         Roger Holt  8/17/1988    M
## 4           4   Suleka Manzanares 1/16/1983    F
## 5           5     Klarissa Winn 11/22/1962    F
## 6           6  Parismonte Gustafson 3/30/1982    M
##           race      street_address      city state zip_code
## 1         White    1906 Kooter Lane  Charlotte   NC   28202
## 2         White 3803 McDonald Avenue Kissimmee   FL   34741
## 3         White   283 Carson Street  Lexington   KY   40517
## 4         White    540 Rafe Lane  Greenwood   MS   38930
## 5 Black/African-American 1693 Simpson Avenue      York   PA   17403
## 6         White    4481 Sunrise Road Las Vegas   NV   89102
##   contact_number hosp_admission hosp_discharge admitting_provider
## 1    704-452-2174   9/15/2012    9/21/2012    Reece E Finch
## 2    407-935-4966   2/19/2012    4/9/2012    Robert C Kaestner
## 3    859-245-2857   2/28/2011    3/20/2011    Chung X Y\xcd_an
## 4    662-707-1035   3/23/2010    4/8/2010    Christian M Burger
```

```
## 5 717-940-9443 1/1/2009 3/26/2009 Archie S Paton
## 6 702-889-8988 3/11/2010 6/4/2010 Ju O Chuang
## had_cardiac_arrests
## 1 1
## 2 0
## 3 0
## 4 0
## 5 0
## 6 0
```

or

```
# Last few rows
tail(d.in)
```

```
## patient.ids patient.names dob gender
## 24102 24114 Nola Castaneda 11/30/1935 F
## 24103 24115 Marlana Sanchez Lopez 11/10/1976 F
## 24104 24116 Anthony Wilson 7/17/1992 M
## 24105 24117 Eugene Lanier 8/11/1948 M
## 24106 24118 Charlene Rice 4/14/1984 F
## 24107 24119 Dominique Johnson 1/21/1961 M
## race street_address city state
## 24102 Asian/Mideast Indian 2427 Maple Lane Huntsville AL
## 24103 White 4836 Park Avenue Sacramento CA
## 24104 White 2667 Sand Fork Road South Bend IN
## 24105 Patient Declined 1772 Scheuvront Drive Denver CO
## 24106 Black/African-American 3322 Medical Center Drive Fort Myers FL
## 24107 Black/African-American 1368 Southern Street Levittown NY
## zip_code contact_number hosp_admission hosp_discharge
## 24102 35816 256-531-9016 7/10/2012 7/19/2012
## 24103 95817 916-450-8202 10/7/2009 10/11/2009
## 24104 46625 574-914-9379 11/19/2010 12/17/2010
## 24105 80216 303-523-3089 2/28/2009 5/1/2009
## 24106 33912 941-621-7330 8/23/2011 9/4/2011
## 24107 11756 516-648-3775 4/1/2010 5/4/2010
## admitting_provider had_cardiac_arrests
## 24102 Nino E Piazza 1
## 24103 Ning B Ting 1
## 24104 Carol E Rowe 0
## 24105 Ryan A Longstaff 0
## 24106 Charles E Fanning 0
## 24107 Thorsten A Beich 0
```

Accessing data frame elements is done as follows

```
# To get a single element in 4th row 7th column
d.in[4,7] # always [row, column]
```

```
## [1] Greenwood
## 4461 Levels: Abbeville Abbyville Abercrombie Aberdeen Abilene ... Zephyrhills
```

```
# To get the 6th element in patient ids
d.in$patient.ids[6]
```

```
## [1] 6
```

```
# or DOB in 7th row
```

```
d.in$dob[7]
```

```
## [1] 6/24/1929
```

```
## 15287 Levels: 1/1/1931 1/1/1932 1/1/1934 1/1/1935 1/1/1936 ... 9/9/1991
```

If I wanted to choose the top 10 rows of this data frame, I would do:

```
d.top10 <- d.in[1:10, ]
```

```
# To get an entire column, use
```

```
d.top10$patient.names
```

```
## [1] Matthew Connor      Joanna Lujan      Roger Holt
```

```
## [4] Suleka Manzanares      Klarissa Winn     Parismonte Gustafson
```

```
## [7] Matthew Gayle          Sophie Vidal Castro Brandon Rishsefid
```

```
## [10] Nicole Carbajal
```

```
## 23689 Levels: A' Kia Torres-Perez A'Jonte Nghiem ... Zyre Sedillos
```

```
# To get the second column
```

```
d.top10[,2]
```

```
## [1] Matthew Connor      Joanna Lujan      Roger Holt
```

```
## [4] Suleka Manzanares      Klarissa Winn     Parismonte Gustafson
```

```
## [7] Matthew Gayle          Sophie Vidal Castro Brandon Rishsefid
```

```
## [10] Nicole Carbajal
```

```
## 23689 Levels: A' Kia Torres-Perez A'Jonte Nghiem ... Zyre Sedillos
```

```
# To get the fifth row
```

```
d.top10[5,]
```

```
## patient.ids patient.names      dob gender      race
```

```
## 5          5 Klarissa Winn 11/22/1962      F Black/African-American
```

```
##      street_address city state zip_code contact_number hosp_admission
```

```
## 5 1693 Simpson Avenue York    PA    17403    717-940-9443    1/1/2009
```

```
## hosp_discharge admitting_provider had_cardiac_arrests
```

```
## 5      3/26/2009      Archie S Paton          0
```

Some useful commands are

```
nrow(d.in) # for number of rows in a data frame
```

```
## [1] 24107
```

```
ncol(d.in) # for number of columns in a data frame
```

```
## [1] 14
```