Introduction to R programming, part 1

Milan Geybels
Steen Ladelund
Majken Hamann Sey

November 27, 2018

1. Simple expressions

```
4
```

```
## [1] 4
```

```
3+4
```

```
## [1] 7
```

```
3^2
```

```
## [1] 9
```

```
1:5
```

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

```
5:1
```

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

```
c(9,3,6)
```

```
## [1] 9 3 6
```

```
?c
```

```
a <- c(9,3,6)
length(a)
```

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

```
a[2]
```

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

```
a+1
```

```
## [1] 10  4  7
```

```
c(a,1)
```

```
## [1] 9 3 6 1
```

Use == to perform a logical test. Can you also test, in one step, if multiple values exist in a? Tip: value matching.

```
a==3
```

```
## [1] FALSE  TRUE FALSE
```

```
a!=3
```

```
## [1]  TRUE FALSE  TRUE
```

```r
sample(1:5)
```

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

```r
print("hello")
```

```
## [1] "hello"
```

```r
c("hi", "there")
```

```
## [1] "hi"    "there"
```

```r
'Hi'=='hi'
```

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

2. Loops and functions

```r
for(i in 1:10){
  sample(1:5)
}
```

What is missing in this code?

```r
pow <- function(x){
  result <- x^2
  paste(x, 'to the power 2 is', result)
}
```

Note that after running this code, the function is part of your `R` environment. Hence, you only need to run the code once.

```r
pow(5)
```

```
## [1] "5 to the power 2 is 25"
```

Can you change the code so that `pow()` can raise a base number to any power?

3. The data frame

The `R` built-in data-set `airquality` is used. You can type `?airquality` to get a description of the data.

```r
air <- airquality
class(air)
```

```
## [1] "data.frame"
```

```r
dim(air)
```

```
## [1] 153   6
```

```r
colnames(air)
```

```
## [1] "Ozone"   "Solar.R" "Wind"    "Temp"    "Month"   "Day"
```

```r
head(air)
```

```
##   Ozone Solar.R Wind Temp Month Day
## 1    41     190  7.4   67     5   1
```

```
## 2    36     118  8.0   72     5   2
## 3    12     149 12.6   74     5   3
## 4    18     313 11.5   62     5   4
## 5    NA      NA 14.3   56     5   5
## 6    28      NA 14.9   66     5   6
```

How are missing values represented?

4. Write and read data

```
getwd()
```

```
## [1] "/Users/milangeybels/Documents/R/R_course"
```

```
write.csv(air, "my_air_data.csv")
air2 <- read.csv("my_air_data.csv", row.names=1)
```

5. Subsetting, the dollar sign, and sorting

```
air[1:6,]
```

```
##   Ozone Solar.R Wind Temp Month Day
## 1    41     190  7.4   67     5   1
## 2    36     118  8.0   72     5   2
## 3    12     149 12.6   74     5   3
## 4    18     313 11.5   62     5   4
## 5    NA      NA 14.3   56     5   5
## 6    28      NA 14.9   66     5   6
```

```
air[1:6, c(2,4)]
```

```
##   Solar.R Temp
## 1     190   67
## 2     118   72
## 3     149   74
## 4     313   62
## 5      NA   56
## 6      NA   66
```

```
air$Ozone
```

```
##   [1]  41  36  12  18  NA  28  23  19   8  NA   7  16  11  14  18  14  34
##  [18]   6  30  11   1  11   4  32  NA  NA  NA  23  45 115  37  NA  NA  NA
##  [35]  NA  NA  NA  29  NA  71  39  NA  NA  23  NA  NA  21  37  20  12  13
##  [52]  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA 135  49  32  NA  64  40  77
##  [69]  97  97  85  NA  10  27  NA   7  48  35  61  79  63  16  NA  NA  80
##  [86] 108  20  52  82  50  64  59  39   9  16  78  35  66 122  89 110  NA
## [103]  NA  44  28  65  NA  22  59  23  31  44  21   9  NA  45 168  73  NA
## [120]  76 118  84  85  96  78  73  91  47  32  20  23  21  24  44  21  28
## [137]   9  13  46  18  13  24  16  13  23  36   7  14  30  NA  14  18  20
```

```
air$Ozone[1:6]
```

```
## [1] 41 36 12 18 NA 28
```

Subsetting can also be done using `TRUE` and `FALSE`.

```r
table(air$Month)
```

```
##
##  5  6  7  8  9
## 31 30 31 31 30
```

```r
august <- air[air$Month==8,]
dim(august)
```

```
## [1] 31  6
```

```r
head(august)
```

```
##    Ozone Solar.R Wind Temp Month Day
## 93    39      83  6.9   81     8   1
## 94     9      24 13.8   81     8   2
## 95    16      77  7.4   82     8   3
## 96    78      NA  6.9   86     8   4
## 97    35      NA  7.4   85     8   5
## 98    66      NA  4.6   87     8   6
```

```r
mean(c(1,2))
```

```
## [1] 1.5
```

```r
mean(air$Ozone)
```

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

```r
is.na(air$Ozone)
```

```
##   [1] FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE  TRUE FALSE
##  [12] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [23] FALSE FALSE  TRUE  TRUE  TRUE FALSE FALSE FALSE FALSE  TRUE  TRUE
##  [34]  TRUE  TRUE  TRUE  TRUE FALSE  TRUE FALSE FALSE  TRUE  TRUE FALSE
##  [45]  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE  TRUE  TRUE
##  [56]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE FALSE FALSE  TRUE FALSE
##  [67] FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE  TRUE FALSE FALSE
##  [78] FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE
##  [89] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [100] FALSE FALSE  TRUE  TRUE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE
## [111] FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE  TRUE FALSE FALSE
## [122] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [133] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [144] FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE
```

```r
air$Ozone[!is.na(air$Ozone)]
```

```
##   [1]  41  36  12  18  28  23  19   8   7  16  11  14  18  14  34   6  30
##  [18]  11   1  11   4  32  23  45 115  37  29  71  39  23  21  37  20  12
##  [35]  13 135  49  32  64  40  77  97  97  85  10  27   7  48  35  61  79
##  [52]  63  16  80 108  20  52  82  50  64  59  39   9  16  78  35  66 122
##  [69]  89 110  44  28  65  22  59  23  31  44  21   9  45 168  73  76 118
##  [86]  84  85  96  78  73  91  47  32  20  23  21  24  44  21  28   9  13
## [103]  46  18  13  24  16  13  23  36   7  14  30  14  18  20
```

```r
mean(air$Ozone[!is.na(air$Ozone)])
```

```
## [1] 42.12931
```

```
mean(air$Ozone, na.rm = TRUE)
```

```
## [1] 42.12931
```

For the next exercise, first generate a data frame named `highO3` that only includes observations with a high ozone concentration of 106 ppb or more (classified as 'Very Unhealthy' by the EPA).

```
highO3
```

```
##     Ozone Solar.R Wind Temp Month Day
## 30    115     223  5.7   79     5  30
## 62    135     269  4.1   84     7   1
## 86    108     223  8.0   85     7  25
## 99    122     255  4.0   89     8   7
## 101   110     207  8.0   90     8   9
## 117   168     238  3.4   81     8  25
## 121   118     225  2.3   94     8  29
```

```
sort(highO3$Ozone)
```

```
## [1] 108 110 115 118 122 135 168
```

```
order(highO3$Ozone)
```

```
## [1] 3 5 1 7 4 2 6
```

```
highO3[order(highO3$Ozone),]
```

```
##     Ozone Solar.R Wind Temp Month Day
## 86    108     223  8.0   85     7  25
## 101   110     207  8.0   90     8   9
## 30    115     223  5.7   79     5  30
## 121   118     225  2.3   94     8  29
## 99    122     255  4.0   89     8   7
## 62    135     269  4.1   84     7   1
## 117   168     238  3.4   81     8  25
```

Can you sort these data by decreasing `Ozone`?