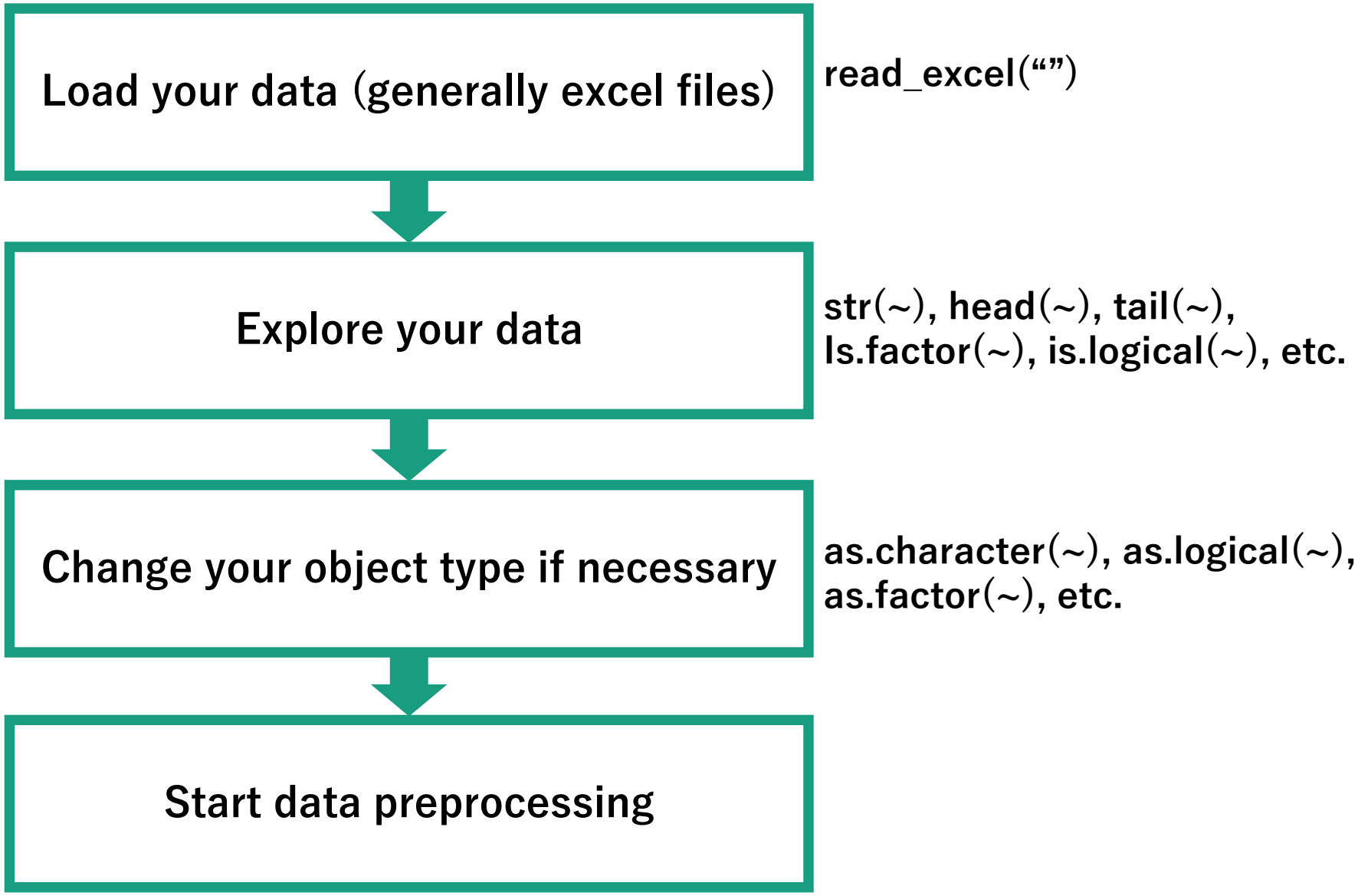
R for toxicology

2. Data preprocessing 1

Objective

- What should we do before we start data preprocessing?

To-do list before preprocessing



Load your data (excel)

Use readxl package to load your data

```
library(readxl)
```

Before you use readxl, you must load it first

You can load them using ‘library(readxl)’

Load your data (excel)

- Use read_excel("~~") from readxl package

```
df <- read_excel("./Example data.xlsx")
head(df)
```

Working directory is hidden in '.' part

```
getwd()
```

You can check working directory by using this

```
setwd("D:/XXXXX")
```

If you want to change working directory, use 'setwd'

Load your data (excel)

Check if your data is loaded without problems

```
> df <- read_excel("./Example data.xlsx")
> head(df)
# A tibble: 6 x 6
  Level Name   `Attack type`    HP   ATK   DEF
  <dbl> <chr>  <chr>         <dbl> <dbl> <dbl>
1    47 Hans   Fire             76    45    22
2    24 Choi   Wind             54    33    12
3    86 Yamaoka Ice             88    64    45
4    78 John   Wind             70    45    23
5    50 Ivan   Earth            92    23    50
6    47 Liu    Fire             74    43    26
```

Explore your data

- Use str(~), head(~), and/or tail(~) functions to briefly explore your dataset

Basic information of the loaded data

```
> str(df)
tibble [10 x 6] (S3: tbl_df/tbl/data.frame)
 $ Level      : num [1:10] 47 24 86 78 50 47 62 52 90 100
 $ Name       : chr [1:10] "Hans" "Choi" "Yamaoka" "John" ...
 $ Attack type: chr [1:10] "Fire" "Wind" "Ice" "Wind" ...
 $ HP         : num [1:10] 76 54 88 70 92 74 63 67 86 94
 $ ATK        : num [1:10] 45 33 64 45 23 43 45 25 85 90
 $ DEF        : num [1:10] 22 12 45 23 50 26 33 68 38 50
```

Explore your data

Show some of your data from the top

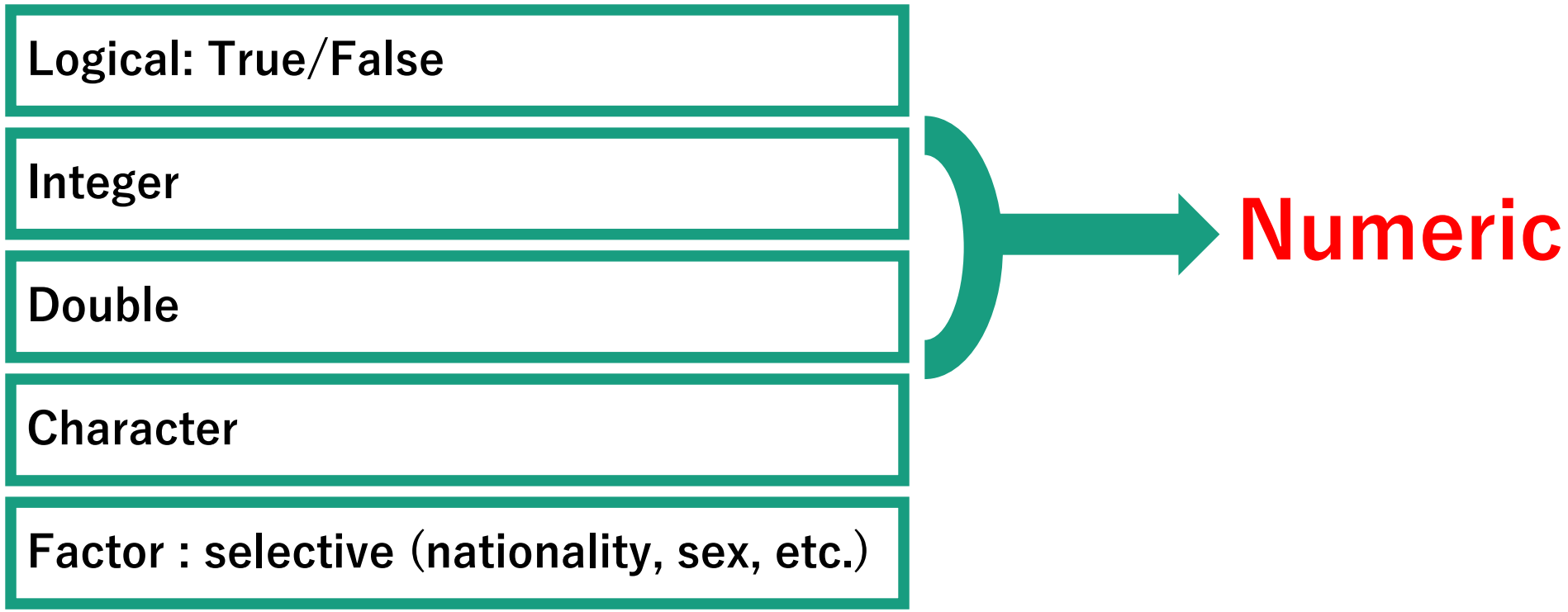
```
> head(df)
# A tibble: 6 x 6
  Level Name    `Attack type`    HP   ATK   DEF
  <dbl> <chr>   <chr>         <dbl> <dbl> <dbl>
1    47 Hans    Fire             76    45    22
2    24 Choi    Wind             54    33    12
3    86 Yamaoka Ice              88    64    45
4    78 John    Wind             70    45    23
5    50 Ivan    Earth            92    23    50
6    47 Liu     Fire             74    43    26
```

Show some of data from the bottom

```
> tail(df)
# A tibble: 6 x 6
  Level Name   `Attack type`    HP   ATK   DEF
  <dbl> <chr>  <chr>         <dbl> <dbl> <dbl>
1    50 Ivan   Earth            92    23    50
2    47 Liu    Fire             74    43    26
3    62 Miguel Ice              63    45    33
4    52 Andres Earth            67    25    68
5    90 Park   Ice              86    85    38
6   100 Ikeda  Fire             94    90    50
```

Explore your data

Five elements that are frequently used in R



Explore your data

- Use class(~$~) to check class of selected objects

```
> class(df$Level)
[1] "numeric"
```

- +)sapply can be applied to explore your dataset at once

```
> sapply(df, class)
     Level       Name Attack type         HP        ATK        DEF
 "numeric" "character" "character"  "numeric"  "numeric"  "numeric"
```

Change your object type

- as.~function is used when you change object type

```
> df
# A tibble: 10 x 6
   Level Name    `Attack type`    HP   ATK   DEF
   <dbl> <chr>   <chr>         <dbl> <dbl> <dbl>
 1    47 Hans    Fire             76    45    22
 2    24 Choi    Wind             54    33    12
 3    86 Yamaoka Ice              88    64    45
 4    78 John    Wind             70    45    23
 5    50 Ivan    Earth            92    23    50
 6    47 Liu     Fire             74    43    26
 7    62 Miguel  Ice              63    45    33
 8    52 Andres  Earth            67    25    68
 9    90 Park    Ice              86    85    38
10   100 Ikeda   Fire             94    90    50
```

This column was changed to factor type because these data are all selective

Change your object type

Change object type of ‘Attack type’ to factor and apply to ‘Attack type’ column

```
> df$`Attack type` <- as.factor(df$`Attack type`)
> is.factor(df$`Attack type`)
[1] TRUE
```

Use is.factor() to check if the objects were successfully changed

You can also use other functions such as as.numeric, as.logical