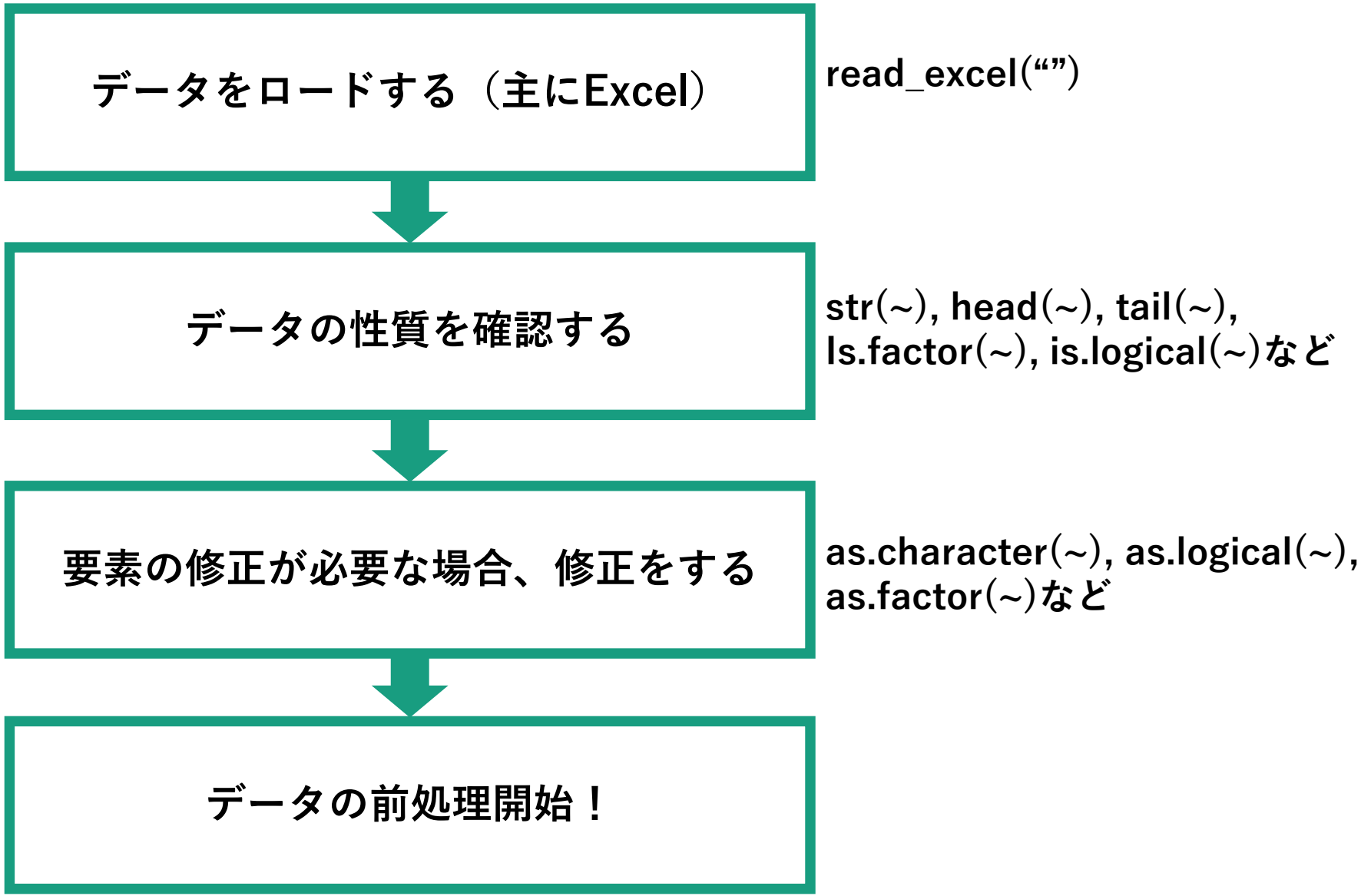
毒性学メンバーのためのR分析

２．データの全処理（１）

主題

- データの前処理に移る前に何をするのか

データの前処理に移る前に



データをロードする（Excel）

readxlパッケージをロードする

```
library(readxl)
```

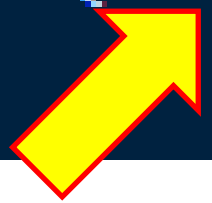
先ずは必要なreadxlパッケージをロードする

パッケージのロードはlibraryでできる

データをロードする（Excel）

- readxlパッケージのread_excel(“~~”)を使う

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



ワーキングディレクトリが隠されている
現在のワーキングディレクトリは

```
getwd()
```

で確認できる

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

ワーキングディレクトリの変更はsetwd関数でできる

データをロードする（Excel）

問題なくロードできたのか確認する

```
> 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
```

データの性質を確認する

- str(~), head(~), tail(~)関数でデータの概略を見る

データの基本的な情報をみる

```
> 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
```

データの性質を確認する

最初のいくつかのデータを見る

```
> 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
```

最後のいくつかのデータを見る

```
> 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
```

データの性質を確認する

頻繁に使うelementの種類五つ

- Logical : True/False二つが存在
- Integer : 整数
- Double : 実数

→ Numeric

- Character : 文字
- Factor : 選択型（例：国籍、性別など）

データの性質を確認する

class(~$~)で要素の種類をチェックする

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

参考)sapply関数を使うと一気に見ることもできる

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

要素の種類を修正をする

as.~~関数を使って要素の種類を変更する

```
> 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
```

この列は、選択肢があるFactorのため、
CharacterではなくFactorに変更する

要素の種類を修正をする

dfのType列をFactorに変換してdfのType列に適用する

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

is.factor()関数を使ってタイプがFactorになったのか確認する。

as.factor以外にもas.numeric, as.logicalなどもある