Data Transformation in R

Kursus R: Pengenalan dan Praktikal (Sesi 2)

Mohd Azmi Bin Suliman

azmi.suliman@moh.gov.my

Pusat Penyelidikan Penyakit Tak Berjangkit, Institut Kesihatan Umum

Wednesday, 02 October 2024

Data Transformation

- Number
- Factor
- Conditional
- Join and Reshape

Setup: Data Transformation

- Create New Level 1 Header
 - → # Data Transformation

Number Transformation

- We can perform basic numerical transformation using R
 - → Don't forget to use mutate function to create new variable
- For example, convert height from cm to m
 - → Input: Height (Height)
 - → Output: Height in meter (Ht_m)
 - → Overwrite current dataset
 - → Hint: use .before or .after parameter to arrange the new variable

```
1 asthmads_spss <- asthmads_spss %>%
2 mutate(Ht_m = Height/100, .after = "Height")
```

For example, convert height in cm (Height) cm to height in m (Ht_m)

```
asthmads spss <- asthmads spss %>%
      mutate(Ht m = Height/100, .after = "Height")
    asthmads spss
\# A tibble: 150 \times 19
                    Age WorkStatus Height Ht m Weight Pre WC Pre PA HW
            Gender
     id idR
  <dbl> <chr> <fct> <dbl> <fct>
                                 <dbl> <dbl>
                                            <dbl> <dbl> <dbl>
     1 ejGs Female 34 Unemployed 179 1.79
                                            84.2
                                                         77
     2 A4pG Male 31 Unemployed
                                            81.8 94
                                  169 1.69
                                            88.5 95
     3 gkCO Male 25 Employed
                                   164 1.64
                                                 53.2 85
     4 jcFZ Female 33 Unemployed
                                   136 1.36
     5 qVSA Male 28 Unemployed
                                               71.3 90
                                   172 1.72
                  33 Unemployed
                                                 87.3 92
                                   178 1.78
     6 wDAR Male
     7 FuAU Female 31 Unemployed
                                                48.8 80
                                   140 1.4
     8 fnKz Female 34 Employed
                                                       82
                                   140 1.4
                                                 49.1
     9 OYTi Male 31 Employed
                                  171 1.71
                                                 60.1
                                                        85
                    28 Employed
10
    10 pfMa Male
                                   163 1.63
                                                 93.1
                                                        101
# i 140 more rows
# i 9 more variables: Weight Post <dbl>, WC Post <dbl>, Tx2 <fct>,
   PEFR Pre <dbl>, PEFR Post <dbl>, SxWheeze Pre <fct>, SxWheeze Post <fct>,
   PS Pre <dbl>, PS Post <dbl>
```

- Now try yourself!
 - → Calculate BMI_Pre and BMI_Post
 - → **Hint**: Use %>% pipe operator to chain the transformation

- Now try yourself!
 - → Calculate BMI_Pre and BMI_Post

```
asthmads spss <- asthmads spss %>%
      mutate(BMI Pre = Weight Pre/(Ht m^2), .after = "Weight Pre") %>%
      mutate(BMI Post = Weight Post/(Ht m^2), .after = "Weight Post")
  4
    asthmads spss
# A tibble: 150 × 21
     id idR
              Gender
                      Age WorkStatus Height Ht m Weight Pre BMI Pre WC Pre
  <dbl> <chr> <fct> <dbl> <fct>
                                                      <dbl>
                                                             <dbl>
                                     <dbl> <dbl>
                                                                   <dbl>
                                                              26.3
      1 ejGs Female
                       34 Unemployed
                                       179
                                           1.79
                                                       84.2
                                                                       77
      2 A4pG Male
                                                            28.6
                      31 Unemployed
                                       169 1.69
                                                      81.8
                                                                       94
                       25 Employed
                                                            32.9
      3 qkCO
             Male
                                       164 1.64
                                                      88.5
                                                                       95
      4 jcFZ
             Female
                       33 Unemployed
                                       136 1.36
                                                       53.2
                                                            28.8
                                                                       85
      5 qVSA
             Male
                       28 Unemployed
                                       172 1.72
                                                      71.3
                                                            24.1
                                                                       90
                   33 Unemployed
      6 wdar
             Male
                                       178 1.78
                                                      87.3
                                                            27.6
                                                                       92
             Female
                       31 Unemployed
                                       140 1.4
                                                       48.8
                                                            24.9
                                                                       80
      7 FuAU
      8 fnKz
             Female
                       34 Employed
                                       140 1.4
                                                            25.1
                                                                       82
                                                       49.1
      9 OYTi
             Male
                       31 Employed
                                       171 1.71
                                                       60.1
                                                              20.6
                                                                       85
                       28 Employed
10
     10 pfMa
             Male
                                       163 1.63
                                                       93.1
                                                              35.0
                                                                      101
# i 140 more rows
# i 11 more variables: PA HW <dbl>, Weight Post <dbl>, BMI Post <dbl>,
                                     Data Transformation in R
```

- We can round the numerical variable using round function
 - → We can round to nearest decimal point by specifying digits parameter

```
asthmads spss <- asthmads spss %>%
       mutate(BMI Pre = round(BMI Pre, digits = 2))
  3
     asthmads spss
# A tibble: 150 × 21
      id idR
                        Age WorkStatus Height Ht m Weight Pre BMI Pre WC Pre
               Gender
   <dbl> <chr> <fct> <dbl> <fct>
                                         <dbl> <dbl>
                                                           <dbl>
                                                                   <dbl>
                                                                          <dbl>
                                                                    26.3
                         34 Unemployed
                                           179
                                                            84.2
                                                                             77
       1 ejGs Female
                                                1.79
                                                                    28.6
       2 A4pG Male
                         31 Unemployed
                                           169 1.69
                                                            81.8
                                                                              94
       3 qkCO
               Male
                         25 Employed
                                           164
                                               1.64
                                                            88.5
                                                                  32.9
                                                                             95
       4 jcFZ
               Female
                         33 Unemployed
                                           136
                                               1.36
                                                            53.2
                                                                    28.8
                                                                             85
                         28 Unemployed
       5 qVSA
                                           172
                                               1.72
                                                            71.3
                                                                    24.1
                                                                              90
               Male
       6 wdar
               Male
                         33 Unemployed
                                           178
                                               1.78
                                                            87.3
                                                                    27.6
                                                                             92
       7 FuAU
               Female
                         31 Unemployed
                                           140
                                                1.4
                                                            48.8
                                                                    24.9
                                                                             80
       8 fnKz
               Female
                         34 Employed
                                           140
                                                            49.1
                                                                    25.0
                                                                             82
                                                1.4
                          31 Employed
                                                                    20.6
       9 OYTi
               Male
                                           171
                                                            60.1
                                                                             85
                                                1.71
10
      10 pfMa
                         28 Employed
                                           163 1.63
                                                            93.1
                                                                    35.0
               Male
                                                                            101
# i 140 more rows
# i 11 more variables: PA_HW <dbl>, Weight Post <dbl>, BMI_Post <dbl>, Data Transformation in R
```

- Now try yourself!
 - → Round BMI_Post to 2 decimal points

- Now try yourself!
 - → Round BMI_Post to 2 decimal points

```
asthmads spss <- asthmads spss %>%
       mutate(BMI Post = round(BMI Post, digits = 2))
  3
     asthmads spss
# A tibble: 150 × 21
      id idR
                        Age WorkStatus Height Ht m Weight Pre BMI Pre WC Pre
               Gender
   <dbl> <chr> <fct> <dbl> <fct>
                                                         <dbl>
                                                                 <dbl>
                                        <dbl> <dbl>
                                                                        <dbl>
                        34 Unemployed
                                         179
                                              1.79
                                                          84.2
                                                                 26.3
                                                                           77
       1 ejGs Female
       2 A4pG Male
                        31 Unemployed
                                         169 1.69
                                                          81.8
                                                                28.6
                                                                           94
                        25 Employed
                                                                32.9
                                                                           95
 3
       3 qkCO Male
                                         164 1.64
                                                          88.5
                                             1.36
                        33 Unemployed
                                         136
                                                          53.2
                                                                28.8
                                                                           85
       4 jcFZ
              Female
       5 qVSA
                        28 Unemployed
                                              1.72
 5
              Male
                                         172
                                                          71.3
                                                                  24.1
                                                                           90
                        33 Unemployed
 6
       6 wdar
              Male
                                         178
                                             1.78
                                                         87.3
                                                                 27.6
                                                                           92
              Female
                        31 Unemployed
                                         140
                                              1.4
                                                          48.8
                                                                 24.9
                                                                           80
       7 FuAU
                        34 Employed
                                                                           82
 8
       8 fnKz
              Female
                                         140
                                              1.4
                                                          49.1
                                                                 25.0
                        31 Employed
       9 OYTi
              Male
                                         171
                                              1.71
                                                          60.1
                                                                 20.6
                                                                           85
10
                         28 Employed
                                                                          101
      10 pfMa
              Male
                                         163 1.63
                                                          93.1
                                                                  35.0
# i 140 more rows
# i 11 more variables: PA HW <dbl>, Weight Post <dbl>, BMI Post <dbl>,
   WC Post <dbl>, Tx2 <fct>, PEFR Pre <dbl>, PEFR Post <dbl>,
    SxWheeze_Pre <fct>, SxWheeze_Post <fct>, PS_Pre <dbl>, PS_Post <dbl>
```

- Please note that in R, rounding is done by to the nearest even number
 - → aka Banker's rounding

```
1 round(1.5)
[1] 2
[1] 1 round(2.5)
[1] 2
```

Practical: Binning (Categorizing Numerical Variable)

- We use cut function to categorize numerical variable
 - → For example, we want to categorize BMI_Pre into
 - → Underweight (< 18.5)
 - → Normal (18.5 22.9)
 - → Overweight (23 24.9)
 - → Obese (> 25)
 - → We use breaks parameter to specify the cut-off points

Practical: Binning (Categorizing Numerical Variable)

 Note: We need to specified small number for the first and large number for the last category

```
asthmads spss <- asthmads spss %>%
      mutate (BMI PreCat = cut (BMI Pre,
                               breaks = c(0, 18.49, 22.99, 24.99, 100)),
              .after = "BMI Pre")
  5
    asthmads spss
# A tibble: 150 × 22
     id idR
                     Age WorkStatus Height Ht m Weight Pre BMI Pre
             Gender
  <dbl> <chr> <fct> <dbl> <fct>
                                   <dbl> <dbl>
                                                  <dbl>
                                                         <dbl>
                                                       26.3
                   34 Unemployed
                                    179 1.79
                                                  84.2
      1 ejGs Female
      2 A4pG Male 31 Unemployed
                                                81.8 28.6
                                    169 1.69
                                                88.5 32.9
      3 gkCO Male 25 Employed
                                    164 1.64
                                                   53.2 28.8
      4 jcFZ Female 33 Unemployed
                                    136 1.36
      5 qVSA Male 28 Unemployed
                                    172 1.72
                                                  71.3
                                                        24.1
      6 wDAR Male 33 Unemployed
                                    178 1.78
                                                  87.3
                                                        27.6
     7 FuAU Female
                    31 Unemployed
                                    140 1.4
                                                  48.8
                                                        24.9
      8 fnKz Female
                    34 Employed
                                    140 1.4
                                                  49.1
                                                        25.0
                  31 Employed
      9 OYTi
            Male
                                    171 1.71
                                              60.1
                                                         20.6
                     28 Employed
                                    163 1.63
     10 pfMa
10
            Male
                                                   93.1
                                                          35.0
# i 140 more rows
# i 13 more variables: BMI_PreCat <fct>, DWG_Pre_SdblationPA_HW <dbl>,
   Weight Post <dbl>, BMI Post <dbl>, WC Post <dbl>, Tx2 <fct>,
```

PEFR_Pre <dbl>, PEFR_Post <dbl>, SxWheeze_Pre <fct>, SxWheeze_Post <fct>,
PS_Pre <dbl>, PS_Post <dbl>

Practical: Binning (Categorizing Numerical Variable)

- We use cut function to categorize numerical variable
 - → For example, we want to categorize BMI_Pre into
 - → Underweight (< 18.5)
 - → Normal (18.5 22.9)
 - → Overweight (23 24.9)
 - → Obese (> 25)
 - → We use labels parameter to label the bin.

Practical: Binning (Categorizing Numerical Variable)

Note the labels = should be one less than breaks =

```
asthmads spss <- asthmads spss %>%
                    mutate (BMI PreCat = cut (BMI Pre,
                                                                                                   breaks = c(0, 18.49, 22.99, 24.99, 100),
                                                                                                   labels = c("Underweight", "Normal", "Overweight", "Ol
      4
                                            .after = "BMI Pre")
      6
             asthmads spss
# A tibble: 150 × 22
                id idR
                                                                  Age WorkStatus Height Ht m Weight Pre BMI Pre
                                        Gender
        <dbl> <chr> <fct> <dbl> <fct>
                                                                                                             <dbl> <dbl>
                                                                                                                                                            <dbl> <dbl>
                                                                                                                                                                             26.3
                   1 ejGs Female 34 Unemployed 179 1.79
                                                                                                                                                                84.2
                                                                                                                                                           81.8 28.6
                   2 A4pG Male 31 Unemployed
                                                                                                                 169 1.69
                                                               25 Employed
                                                                                                                                                        88.5 32.9
                   3 qkCO Male
                                                                                                                  164 1.64
                                                                                                                                                                53.2 28.8
                   4 jcFZ Female
                                                                33 Unemployed
                                                                                                                  136 1.36
                                                                                                                                                        71.3 24.1
   5
                   5 qVSA Male
                                                             28 Unemployed
                                                                                                                  172 1.72
                                                       33 Unemployed
                                                                                                                                                                                 27.6
                   6 wDAR Male
                                                                                                                  178 1.78
                                                                                                                                                           87.3
                  7 FuAU Female
                                                                 31 Unemployed
                                                                                                                   140 1.4
                                                                                                                                                             48.8
                                                                                                                                                                                 24.9
                                                                 34 Employed
                                                                                                                                                                                 25.0
                  8 fnKz Female
                                                                                                                   140 1.4
                                                                                                                                                             49.1
                   9 OYTi
                                        Male
                                                                    31 Employed
                                                                                                                   171 1.71
                                                                                                                                                                60.1
                                                                                                                                                                                 20.6
                                                                    28 Employed
                10 pfMa
                                        Male
                                                                                                                   163 1.63
                                                                                                                                                                93.1
                                                                                                                                                                                      35.0
10
# i 140 more rows
# i 13 more variables: BMI PreCat <fct>, WC Pre <dbl>, PA HW <dbl>,
          Weight Post <dbl>, BMI Post <dbl>, weather an attention of the rest of the res
```

Practical: Binning

- Now try yourself!
 - → Categorized BMI_Post into the same category as BMI_Pre

Practical: Binning

- Now try yourself!
 - → Categorized BMI_Post into the same category as BMI_Pre

```
asthmads spss <- asthmads spss %>%
       mutate (BMI PostCat = cut (BMI Post,
  3
                                   breaks = c(0, 18.49, 22.99, 24.99, 100),
                                   labels = c("Underweight", "Normal", "Overweight", "O)
  4
  5
               .after = "BMI Post")
  6
    asthmads spss
# A tibble: 150 \times 23
     id idR
                       Age WorkStatus Height Ht m Weight Pre BMI Pre
              Gender
  <dbl> <chr> <fct> <dbl> <fct>
                                       <dbl> <dbl>
                                                       <dbl>
                                                               <dbl>
      1 ejGs
                                        179 1.79
                                                        84.2
                                                                26.3
             Female
                        34 Unemployed
                        31 Unemployed
      2 A4pG Male
                                        169 1.69
                                                        81.8
                                                               28.6
      3 qkCO
                        25 Employed
                                        164 1.64
                                                              32.9
              Male
                                                        88.5
                        33 Unemployed
                                        136 1.36
                                                               28.8
      4 jcFZ
              Female
                                                        53.2
      5 qVSA
                        28 Unemployed
                                        172 1.72
                                                              24.1
              Male
                                                        71.3
      6 wdar
              Male
                        33 Unemployed
                                        178 1.78
                                                        87.3
                                                              27.6
      7 FuAU
              Female
                        31 Unemployed
                                        140 1.4
                                                        48.8
                                                              24.9
                        34 Employed
                                                              25.0
      8 fnKz
              Female
                                        140 1.4
                                                        49.1
      9 OYTi
              Male
                        31 Employed
                                        171 1.71
                                                        60.1
                                                                20.6
     10 pfMa
                        28 Employed
                                                                35.0
              Male
                                      Data Frants for Phation in R93.1
10
# i 140 more rows
```

i 14 more variables: BMI_PreCat <fct>, WC_Pre <dbl>, PA_HW <dbl>,

Weight_Post <dbl>, BMI_Post <dbl>, BMI_PostCat <fct>, WC_Post <dbl>,

Tx2 <fct>, PEFR_Pre <dbl>, PEFR_Post <dbl>, SxWheeze_Pre <fct>,

\$xWheeze_Post <fct>, PS_Pre <dbl>, PS_Post <dbl>

Factor Transformation

Practical: Modifying Factor Order

- We can modify the order of factor using fct relevel function
 - → For example, we want to change the order of Gender
 - → First we check the current order of Gender using levels function

```
levels(asthmads spss$Gender)
```

"Male" "Female"

Practical: Modifying Factor Order

- We can modify the order of factor using fct_relevel function
 - → For example, we want to change the order of Gender
 - → We can either write the full order, or the first order

```
asthmads spss <- asthmads spss %>%
       mutate(Gender F = fct relevel(Gender, "Female", "Male"),
  3
               .after = "Gender")
  4
    asthmads spss
# A tibble: 150 \times 24
              Gender Gender F
                                Age WorkStatus Height Ht m Weight Pre BMI Pre
  <dbl> <chr> <fct> <fct>
                              <dbl> <fct>
                                                <dbl> <dbl>
                                                                <dbl>
                                                                        <dbl>
      1 ejGs Female Female
                                 34 Unemployed
                                                 179 1.79
                                                                 84.2
                                                                         26.3
      2 A4pG
              Male
                                 31 Unemployed
                                                 169 1.69
                                                                 81.8
                                                                         28.6
                     Male
                                                 164 1.64
                                 25 Employed
                                                                 88.5
                                                                         32.9
      3 qkCO
              Male
                     Male
      4 icFZ
              Female Female
                                 33 Unemployed
                                                 136 1.36
                                                                 53.2
                                                                         28.8
                                 28 Unemployed
      5 qVSA
              Male
                     Male
                                                 172 1.72
                                                                 71.3
                                                                         24.1
      6 wdar
              Male
                                 33 Unemployed
                                                 178 1.78
                                                                 87.3
                                                                         27.6
                     Male
              Female Female
                                 31 Unemployed
                                                 140 1.4
                                                                 48.8
                                                                         24.9
      7 FuAU
      8 fnKz
              Female Female
                                 34 Employed
                                                 140 1.4
                                                                 49.1
                                                                         25.0
                                 31 Employed
                                                 171 1.71
                                                                 60.1
                                                                         20.6
      9 OYTi
              Male
                     Male
                                 28 Employed
     10 pfMa
                                                                 93.1
                                                                         35.0
              Male
                     Male
# i 140 more rows
```

i 14 more variables: BMI_PreCat <fct>, WC_Pre <dbl>, PA_HW <dbl>,

Weight_Post <dbl>, BMI_Post <dbl>, BMI_PostCat <fct>, WC_Post <dbl>,

Tx2 <fct>, PEFR_Pre <dbl>, PEFR_Post <dbl>, SxWheeze_Pre <fct>,

\$xWheeze_Post <fct>, PS_Pre <dbl>, PS_Post <dbl>

Practical: Modifying Factor Order

- We can modify the order of factor using fct_relevel function
 - → For example, we want to change the order of Gender
 - → We can either write the full order, or the first order

```
asthmads spss <- asthmads spss %>%
       mutate(Gender F = fct relevel(Gender, "Female"),
  3
               .after = "Gender")
  4
    asthmads spss
# A tibble: 150 \times 24
              Gender Gender F
                                Age WorkStatus Height Ht m Weight Pre BMI Pre
  <dbl> <chr> <fct> <fct>
                              <dbl> <fct>
                                                <dbl> <dbl>
                                                                 <dbl>
                                                                         <dbl>
      1 ejGs Female Female
                                 34 Unemployed
                                                  179 1.79
                                                                  84.2
                                                                         26.3
      2 A4pG
              Male
                                 31 Unemployed
                                                 169 1.69
                                                                  81.8
                                                                         28.6
                     Male
                                                 164 1.64
                                 25 Employed
                                                                 88.5
                                                                         32.9
      3 qkCO
              Male
                     Male
      4 icFZ
              Female Female
                                 33 Unemployed
                                                 136 1.36
                                                                  53.2
                                                                         28.8
                                 28 Unemployed
      5 qVSA
              Male
                     Male
                                                  172 1.72
                                                                  71.3
                                                                         24.1
      6 wdar
              Male
                                 33 Unemployed
                                                  178 1.78
                                                                  87.3
                                                                         27.6
                     Male
              Female Female
                                 31 Unemployed
                                                  140 1.4
                                                                  48.8
                                                                         24.9
      7 FuAU
      8 fnKz
              Female Female
                                 34 Employed
                                                  140 1.4
                                                                  49.1
                                                                         25.0
                                 31 Employed
                                                  171 1.71
                                                                  60.1
                                                                         20.6
      9 OYTi
              Male
                     Male
                                 28 Employed
     10 pfMa
                                                                  93.1
                                                                          35.0
              Male
                     Male
# i 140 more rows
```

i 14 more variables: BMI_PreCat <fct>, WC_Pre <dbl>, PA_HW <dbl>,

Weight_Post <dbl>, BMI_Post <dbl>, BMI_PostCat <fct>, WC_Post <dbl>,

Tx2 <fct>, PEFR_Pre <dbl>, PEFR_Post <dbl>, SxWheeze_Pre <fct>,

\$xWheeze_Post <fct>, PS_Pre <dbl>, PS_Post <dbl>

Practical: Modifying Factor Order

- We can modify the order of factor using fct_relevel function
 - → For example, we want to change the order of Gender

- We can also recode factor using fct_recode function
 - → For example, we want to change Treatment Group (Tx2) to Intervention Group (Tx1)
 - → Placebo to Control
 - → Drug A and Drug B to Intervention
 - → Note: Recode factor does not change the order of the factor

```
asthmads spss <- asthmads spss %>%
 2
      mutate (Tx1 = fct recode (Tx2,
 3
                               "Control" = "Placebo",
 4
                               "Intervention" = "Drug A",
 5
                               "Intervention" = "Drug B"),
 6
              .before = Tx2)
    asthmads spss
# A tibble: 150 × 25
     id idR
            Gender Gender F
                            Age WorkStatus Height Ht m Weight Pre BMI Pre
  <dbl> <chr> <fct> <fct> <dbl> <fct>
                                          <dbl> <dbl>
                                                         <dbl>
                                                                 <dbl>
      1 ejGs Female Female
                             34 Unemployed
                                            179 1.79
                                                          84.2
                                                                 26.3
                                           169 1.69
      2 A4pG Male
                  Male 31 Unemployed
                                                          81.8 28.6
                                                          88.5 32.9
      3 qkCO
            Male
                  Male 25 Employed
                                           164 1.64
                                           136 1.36
      4 jcFZ
            Female Female
                             33 Unemployed
                                                          53.2 28.8
      5 qVSA Male Male 28 Unemployed
                                            172 1.72
                                                          71.3 24.1
                                                          87.3 27.6
      6 wDAR Male Male
                             33 Unemployed
                                            178 1.78
     7 FuAU Female Female 31 Unemployed
                                                          48.8 24.9
                                            140 1.4
                                                          49.1 25.0
                         34 Employed
      8 fnKz Female Female
                                            140 1.4
      9 OYTi Male Male
                             31 Employed
                                            171 1.71
                                                          60.1 20.6
                             28 Employed
10
     10 pfMa Male Male
                                            163 1.63
                                                          93.1
                                                                 35.0
# i 140 more rows
# i 15 more variables: BMI PreCat <fct>, WC Pre <dbl>, PA HW <dbl>,
   Weight Post <dbl>, BMI Post <dbl>, BMI PostCat <fct>, WC Post <dbl>,
   Tx1 <fct>, Tx2 <fct>, PEFR Pre <dbl>, PEFR Post <dbl>, SxWheeze Pre <fct>,
   SxWheeze Post <fct>, PS Pre <dbl>, PSateo Inans for mation in R
```

- We can also recode factor using fct_recode function
 - → can also be use to collapse the factor

```
asthmads spss <- asthmads spss %>%
 2
      mutate (Obese PreCat = fct recode (BMI PreCat,
 3
                                       "Non-Obese" = "Underweight",
                                       "Non-Obese" = "Normal",
 4
 5
                                       "Non-Obese" = "Overweight"),
 6
             .after = "BMI PreCat")
    asthmads spss
# A tibble: 150 \times 26
     id idR
            Gender Gender F Age WorkStatus Height Ht m Weight Pre BMI Pre
  <dbl> <chr> <fct> <fct> <dbl> <fct>
                                         <dbl> <dbl>
                                                        <dbl>
                                                               <dbl>
     1 ejGs Female Female 34 Unemployed
                                          179 1.79
                                                        84.2 26.3
     2 A4pG Male
                 Male 31 Unemployed
                                          169 1.69 81.8 28.6
                                          164 1.64 88.5 32.9
     3 qkCO Male
                 Male 25 Employed
                                          136 1.36 53.2 28.8
     4 jcFZ
            Female Female 33 Unemployed
     5 qVSA Male Male 28 Unemployed
                                          172 1.72
                                                        71.3 24.1
     6 wDAR Male Male 33 Unemployed
                                          178 1.78
                                                        87.3 27.6
     7 FuAU Female Female 31 Unemployed
                                                        48.8 24.9
                                          140 1.4
     8 fnKz Female Female 34 Employed
                                                       49.1 25.0
                                          140 1.4
     9 OYTi Male Male
                          31 Employed
                                          171 1.71 60.1 20.6
                            28 Employed
10
     10 pfMa Male Male
                                           163 1.63
                                                         93.1 35.0
# i 140 more rows
# i 16 more variables: BMI PreCat <fct>, Obese PreCat <fct>, WC Pre <dbl>,
   PA HW <dbl>, Weight Post <dbl>, BMI Post <dbl>, BMI PostCat <fct>,
   WC Post <dbl>, Tx1 <fct>, Tx2 <fct>, PEFR Pre <dbl>, PEFR Post <dbl>,
   SxWheeze Pre <fct>, SxWheeze Post <feta, Transferenationin, R PS Post <dbl>
```

Practical: Collapse Factor

- In previous example, we use fct_recode to collapse the factor
- Collapsing factor have their own function, i.e., fct_collapse

Practical: Collapse Factor

```
asthmads spss <- asthmads spss %>%
 2
      mutate (Obese PostCat = fct collapse (BMI PostCat,
 3
                                       "Non-Obese" = c("Underweight",
 4
                                                       "Normal",
 5
                                                       "Overweight")),
             .after = "BMI PostCat")
    asthmads spss
# A tibble: 150 \times 27
     id idR
            Gender Gender F Age WorkStatus Height Ht m Weight Pre BMI Pre
  <dbl> <chr> <fct> <fct> <dbl> <fct>
                                         <dbl> <dbl>
                                                        <dbl>
                                                               <dbl>
     1 ejGs Female Female 34 Unemployed
                                          179 1.79
                                                         84.2 26.3
     2 A4pG Male
                 Male 31 Unemployed
                                          169 1.69 81.8 28.6
                 Male 25 Employed
                                          164 1.64 88.5 32.9
     3 qkCO
            Male
            Female Female 33 Unemployed
                                          136 1.36
                                                        53.2 28.8
     4 jcFZ
     5 qVSA Male Male 28 Unemployed
                                          172 1.72
                                                        71.3 24.1
     6 wDAR Male Male 33 Unemployed
                                          178 1.78
                                                        87.3 27.6
     7 FuAU Female Female 31 Unemployed
                                                       48.8 24.9
                                          140 1.4
                        34 Employed
                                                        49.1 25.0
     8 fnKz Female Female
                                          140 1.4
     9 OYTi Male Male
                          31 Employed
                                          171 1.71
                                                       60.1 20.6
                            28 Employed
10
     10 pfMa Male Male
                                           163 1.63
                                                         93.1 35.0
# i 140 more rows
# i 17 more variables: BMI PreCat <fct>, Obese PreCat <fct>, WC Pre <dbl>,
   PA HW <dbl>, Weight Post <dbl>, BMI Post <dbl>, BMI PostCat <fct>,
   Obese PostCat <fct>, WC Post <dbl>, Tx1 <fct>, Tx2 <fct>, PEFR Pre <dbl>,
                                  Data Transformation in R
```

Conditional Transformation

Conditional Transformation

- Sometimes, we want to create new variable based on certain condition.
- Commonly, two functions are used for conditional transformation:

```
→ if_else()
→ case_when()
```

Practical: if_else() Conditional Transformation

- For simple conditional transformation, if_else() function is preferable
- if_else() function also were common in other software like Excel, SPSS, etc.

Practical: if_else() Conditional Transformation

```
asthmads spss <- asthmads spss %>%
     mutate(Obese Pre = if else(BMI PreCat == "Obese", "Yes", "No"),
             .after = "BMI PreCat")
 4
    asthmads spss
# A tibble: 150 × 28
    id idR Gender Gender F Age WorkStatus Height Ht m Weight Pre BMI Pre
  <dbl> <chr> <fct> <fct> <dbl> <fct> <dbl> <dbl>
                                                     <dbl> <dbl>
     1 ejGs Female Female 34 Unemployed 179 1.79 84.2
                                                           26.3
     2 A4pG Male Male 31 Unemployed 169 1.69 81.8 28.6
     3 qkCO Male Male 25 Employed 164 1.64 88.5 32.9
     4 jcFZ Female Female 33 Unemployed
                                        136 1.36 53.2 28.8
     5 qVSA Male Male 28 Unemployed
                                        172 1.72 71.3 24.1
     6 wDAR Male Male 33 Unemployed
                                        178 1.78
                                                      87.3 27.6
     7 FuAU Female Female 31 Unemployed
                                                  48.8 24.9
                                         140 1.4
     8 fnKz Female Female 34 Employed
                                                  49.1 25.0
                                        140 1.4
                       31 Employed
                                         171 1.71 60.1 20.6
     9 OYTi Male Male
    10 pfMa Male Male 28 Employed
                                                      93.1 35.0
10
                                         163 1.63
# i 140 more rows
# i 18 more variables: BMI PreCat <fct>, Obese Pre <chr>, Obese PreCat <fct>,
   WC Pre <dbl>, PA HW <dbl>, Weight Post <dbl>, BMI Post <dbl>,
   BMI PostCat <fct>, Obese PostCat <fct>, WC Post <dbl>, Tx1 <fct>,
   Tx2 <fct>, PEFR Pre <dbl>, PEFR Post <dbl>, SxWheeze Pre <fct>,
   SxWheeze Post <fct>, PS Pre <dbl>, PS Post <dbl>
```

Practical: case_when() Conditional Transformation

- For more complex conditional transformation, case_when() function is preferable
 - → For example, nested if_else() in if_else() function
- In this example, we want to categorise abdominal obesity (Abd_Obese), based on this criteria:
 - → If Male and Waist Circumference >= 90 cm: Abdominal Obese
 - → If Male and Waist Circumference < 90 cm: No Abdominal Obese
 - → If Female and Waist Circumference >= 80 cm: Abdominal Obese
 - → If Female and Waist Circumference < 80 cm: No Abdominal Obese</p>

Practical: case_when() Conditional Transformation

```
asthmads spss <- asthmads spss %>%
     mutate(AbdObese Pre = case when(
       WC Pre >= 90 & Gender == "Male" ~ "Abdominal Obese",
       WC Pre < 90 & Gender == "Male" ~ "No Abdominal Obese",
 4
       WC Pre >= 80 & Gender == "Female" ~ "Abdominal Obese",
       WC Pre < 80 & Gender == "Female" ~ "No Abdominal Obese"
       ),
        .after = "WC Pre")
 9
   asthmads spss
# A tibble: 150 × 29
                          Age WorkStatus Height Ht m Weight Pre BMI Pre
    id idR
           Gender Gender F
  <dbl> <chr> <fct> <fct> <dbl> <fct>
                                    <dbl> <dbl>
                                                     <dbl> <dbl>
     1 ejGs Female Female 34 Unemployed 179 1.79 84.2 26.3
     2 A4pG Male Male 31 Unemployed 169 1.69 81.8 28.6
     3 qkCO Male Male 25 Employed 4 jcFZ Female Female 33 Unemployed
                                        164 1.64 88.5 32.9
                                        136 1.36
                                                      53.2 28.8
     5 qVSA Male Male 28 Unemployed
5
                                        172 1.72
                                                      71.3 24.1
     6 wDAR Male Male 33 Unemployed
                                        178 1.78
                                                      87.3 27.6
     7 FuAU Female Female 31 Unemployed
                                                    48.8
                                                           24.9
                                         140 1.4
                                        140 1.4
                                                      49.1 25.0
     8 fnKz Female Female 34 Employed
```

31 Employed 171 1.71 60.1

163 1.63

20.6

35.0

93.1

i 140 more rows # i 19 more variables: BMI_PreCat <fct>, Dobesternation in Obese_PreCat <fct>,

Male 28 Employed

9 OYTi Male Male

10 pfMa Male

10

WC_Pre <dbl>, AbdObese_Pre <chr>, PA_HW <dbl>, Weight_Post <dbl>,
BMI_Post <dbl>, BMI_PostCat <fct>, Obese_PostCat <fct>, WC_Post <dbl>,
Tx1 <fct>, Tx2 <fct>, PEFR_Pre <dbl>, PEFR_Post <dbl>, SxWheeze_Pre <fct>,
SyWhoozo Post <fct>, DS_Pro <dbl>, DS_Post <dbl>

Join and Reshape

- inner_join()
- left_join()
- right_join()
- full_join()
- pivot_longer()
- pivot_wider()

Join Dataset

- Occasionally we have two dataset that correspond to same individual
 - → e.g., pre and post intervention
 - → e.g., demographic and clinical data
- We can join the dataset using *_join() function
 - → inner_join(), left_join(), right_join(), full_join()
 - → matching key is required, e.g., id
- Add another Level 1 Header
 - \rightarrow ## Join Dataset

- In this example, we have two SPSS dataset.
- We need to import it.

```
1 asthmads_pre <- read_sav("Dataset/asthmads_pre.sav") %>%
2 as_factor()
3
4 asthmads_post <- read_sav("Dataset/asthmads_post.sav") %>%
5 as_factor()
```

 In the asthmads_pre dataset, we have baseline measurement and identifier idR

```
asthmads pre
# A tibble: 150 \times 12
     id idR
                      Age WorkStatus Height Weight Pre WC Pre Tx2
                                                                   PEFR Pre
              Gender
  <dbl> <chr> <fct> <dbl> <fct>
                                     <dbl>
                                               <dbl> <dbl> <fct>
                                                                      <dbl>
                                                84.2
                                                                        397
      1 ejGs Female
                       34 Unemployed
                                       179
                                                         77 Placebo
                                                81.8
                  31 Unemployed
      2 A4pG
             Male
                                       169
                                                         94 Placebo
                                                                        472
                                                88.5 95 Placebo
      3 gkCO
                  25 Employed
                                       164
                                                                        476
             Male
                                                53.2 85 Placebo
      4 jcFZ
             Female 33 Unemployed
                                       136
                                                                        416
                                                71.3 90 Drug A
      5 qVSA Male 28 Unemployed
                                       172
                                                                        452
                                                87.3 92 Placebo
                  33 Unemployed
                                       178
      6 wDAR Male
                                                                        484
                                               48.8 80 Drug A
49.1 82 Placebo
      7 FuAU Female 31 Unemployed
                                       140
                                                                        366
      8 fnKz Female 34 Employed
                                       140
                                                                        435
                                             60.1 85 Placebo
 9
      9 OYTi
             Male
                       31 Employed
                                       171
                                                                        425
                       28 Employed
10
     10 pfMa Male
                                       163
                                                93.1
                                                        101 Placebo
                                                                        437
# i 140 more rows
# i 2 more variables: SxWheeze Pre <fct>, PS Pre <dbl>
```

- In the asthmads_post dataset, we have missing baseline measurement
- But we have post measurment and identifier idR

```
asthmads post
# A tibble: 150 \times 5
             PEFR Post SxWheeze Post PS Post
  <dbl> <chr>
                 <dbl> <fct>
                                      <dbl>
      1 ejGs
                   355 No
      2 A4pG
             445 Yes
             481 Yes
      3 qkCO
               382 No
      4 jcFZ
                475 No
      5 qVSA
                497 No
336 No
      6 wDAR
      7 FuAU
                413 No
      8 fnKz
                434 No
      9 OYTi
10
     10 pfMa
               413 No
# i 140 more rows
```

- We can use *_join() function to join the dataset
 - → The function can automatically detect the matching key
 - → However, preferably we specify the matching key using by parameter

```
asthmads join <- left join (asthmads pre, asthmads post, by = "idR")
  2
    asthmads join
# A tibble: 150 × 16
   id.x idR
                       Age WorkStatus Height Weight Pre WC Pre Tx2
                                                                     PEFR Pre
              Gender
                                                 <dbl> <dbl> <fct>
  <dbl> <chr> <fct> <dbl> <fct>
                                      <dbl>
                                                                        <dbl>
      1 ejGs Female
                       34 Unemployed
                                        179
                                                  84.2
                                                          77 Placebo
                                                                          397
                       31 Unemployed
                                        169
                                                                          472
      2 A4pG
             Male
                                                  81.8
                                                          94 Placebo
                       25 Employed
                                        164
      3 qkCO
              Male
                                                 88.5 95 Placebo
                                                                          476
                       33 Unemployed
                                        136
              Female
                                                  53.2
                                                                          416
      4 icFZ
                                                          85 Placebo
      5 qVSA
              Male
                       28 Unemployed
                                        172
                                                  71.3
                                                          90 Drug A
                                                                          452
      6 wDAR
              Male
                       33 Unemployed
                                        178
                                                  87.3
                                                          92 Placebo
                                                                          484
                                                                          366
              Female
                       31 Unemployed
                                        140
                                                 48.8
      7 FuAU
                                                          80 Drug A
                                        140
                                                 49.1
      8 fnKz
              Female
                       34 Employed
                                                          82 Placebo
                                                                          435
      9 OYTi
              Male
                       31 Employed
                                        171
                                                  60.1
                                                          85 Placebo
                                                                          425
10
     10 pfMa
              Male
                       28 Employed
                                        163
                                                  93.1
                                                         101 Placebo
                                                                          437
# i 140 more rows
```

Reshape Dataset

- Sometimes, we want to reshape the dataset
 - → From wide to long
 - → From long to wide
- For example, certain analysis only accept long format
- Add another Level 1 Header
 - → ## Reshape Dataset

Practical: Reshape Dataset

 In this example, we have wide dataset with idR, PEFR_Pre, and PEFR_Post

```
asthma wide <- asthmads join %>%
       select(idR:Tx2, PEFR Pre, PEFR Post)
     asthma wide
# A tibble: 150 \times 10
   idR
                  Age WorkStatus Height Weight Pre WC Pre Tx2
                                                                   PEFR Pre
         Gender
   <chr> <fct> <dbl> <fct>
                                  <dbl>
                                             <dbl> <dbl> <fct>
                                                                      <dbl>
 1 ejGs Female
                   34 Unemployed
                                    179
                                         84.2
                                                        77 Placebo
                                                                        397
                                         81.8 94 Placebo
 2 A4pG Male 31 Unemployed
                                    169
                                                                        472
                                         81.8 94 Placebo
88.5 95 Placebo
53.2 85 Placebo
71.3 90 Drug A
87.3 92 Placebo
48.8 80 Drug A
49.1 82 Placebo
 3 qkCO Male 25 Employed
                                    164
                                                                        476
 4 jcFZ Female 33 Unemployed
                                    136
                                                                        416
 5 qVSA Male 28 Unemployed
                                    172
                                                                        452
                33 Unemployed
 6 wDAR
        Male
                                    178
                                                                        484
 7 FuAU Female 31 Unemployed
                                    140
                                                                        366
 8 fnKz Female 34 Employed
                                    140
                                                                        435
 9 OYTi Male
                31 Employed
                                    171
                                              60.1 85 Placebo
                                                                        425
10 pfMa Male
                   28 Employed
                                    163
                                              93.1 101 Placebo
                                                                        437
# i 140 more rows
# i 1 more variable: PEFR Post <dbl>
```

Practical: Reshape Dataset

- We can use pivot_longer() function to reshape the dataset
 - → We specify the cols parameter to specify the columns to be reshaped
 - → We specify the names_to parameter to specify the new column name for the reshaped columns
 - → We specify the values_to parameter to specify the new column name for the values of the reshaped columns

Practical: Reshape Dataset

```
Age WorkStatus Height Weight Pre WC Pre Tx2 Time
  idR
       Gender
                                                               PEFR
  <chr> <fct> <dbl> <fct>
                                   <dbl> <dbl> <fct> <chr> <dbl>
                             <dbl>
1 ejGs Female
                34 Unemployed
                               179
                                        84.2
                                                                397
                                               77 Placebo Pre
2 ejGs Female 34 Unemployed
                               179
                                   84.2 77 Placebo Post
                                                                355
                                   81.8 94 Placebo Pre
                                                                472
       Male 31 Unemployed
                               169
 3 A4pG
                                   81.8 94 Placebo Post
       Male 31 Unemployed
                               169
                                                                445
4 A4pG
                                       88.5 95 Placebo Pre
            25 Employed
       Male
                               164
                                                                476
 5 qkCO
                                       88.5 95 Placebo Post
 6 qkCO Male 25 Employed
                               164
                                                                481
                                       53.2 85 Placebo Pre
7 jcFZ Female 33 Unemployed
                               136
                                                                416
                                       53.2 85 Placebo Post
                               136
                                                                382
                33 Unemployed
8 jcFZ
       Female
                                       71.3
                               172
                                                                452
9 qVSA
                28 Unemployed
                                               90 Drug A Pre
       Male
                                       71.3
10 qVSA Male
                28 Unemployed
                               172
                                                90 Drug A Post
                                                                475
# i 290 more rows
```