Data wrangling

Steve Huguenin

2025-10-13

Part one - Wrangling

This tutorial will allow you to explore `dplyr` functionality based on the previous lecture. Every question can be answered with a combination of `|>` pipes. You should refrain from using temporary variables and statements outside of the range of the tidyverse.

The first part does not require joins or pivots.

Import the data from the website.

Assign to the name `judgments` and correct the column types where needed.

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.5
v forcats   1.0.1     v stringr   1.5.2
v ggplot2   4.0.0     v tibble    3.3.0
v lubridate 1.9.4     v tidyr     1.3.1
v purrr     1.1.0
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to be
```

```
judgments <- read_tsv("https://biostat2.uni.lu/practicals/data/judgments.tsv")
```

```
Rows: 188 Columns: 158
-- Column specification --------------------------------------------------------
Delimiter: "\t"
chr   (5): start_date, end_date, condition, gender, logbook
dbl (153): finished, subject, age, mood_pre, mood_post, STAI_pre_1_1, STAI_p...
```

```
i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
glimpse(judgments) # Rows: 188 Columns: 158s
```

```
Rows: 188
Columns: 158
$ start_date                <chr> "11/3/2014", "11/3/2014", "11/3/2014", "11/~
$ end_date                  <chr> "11/3/2014", "11/3/2014", "11/3/2014", "11/~
$ finished                  <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1~
$ condition                 <chr> "control", "stress", "stress", "stress", "c~
$ subject                   <dbl> 2, 1, 3, 4, 7, 6, 5, 9, 16, 13, 18, 14, 12,~
$ gender                    <chr> "female", "female", "female", "female", "fe~
$ age                       <dbl> 24, 19, 19, 22, 22, 22, 18, 20, 21, 19, 19,~
$ mood_pre                  <dbl> 81, 59, 22, 53, 48, 73, NA, 100, 67, 30, 55~
$ mood_post                 <dbl> NA, 42, 60, 68, NA, 73, NA, NA, 74, 68, 57,~
$ STAI_pre_1_1              <dbl> 2, 3, 4, 2, 1, 2, 2, 1, 2, 4, 2, 1, 2, 1, 1~
$ STAI_pre_1_2              <dbl> 1, 2, 3, 2, 1, 2, 2, 1, 2, 2, 3, 2, 2, 1, 1~
$ STAI_pre_1_3              <dbl> 2, 3, 3, 2, 1, 1, 1, 1, 1, 3, 1, 2, 2, 2, 2~
$ STAI_pre_1_4              <dbl> 2, 1, 3, 2, 1, 1, 1, 1, 1, 3, 1, 2, 1, 1, 1~
$ STAI_pre_1_5              <dbl> 2, 3, 4, 3, 2, 2, 2, 1, 2, 3, 2, 2, 2, 2, 2~
$ STAI_pre_1_6              <dbl> 2, 2, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 1~
$ STAI_pre_1_7              <dbl> 2, 3, 3, 1, 1, 2, 1, 1, 1, 3, 1, 1, 2, 1, 3~
$ STAI_pre_2_1              <dbl> 2, 3, 4, 3, 3, 2, 2, 2, 2, 4, 3, 3, 2, 4, 3~
$ STAI_pre_2_2              <dbl> 1, 2, 2, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 1~
$ STAI_pre_2_3              <dbl> 1, 2, 3, 3, 3, 2, 2, 1, 2, 3, 2, 3, 3, 3, 2~
$ STAI_pre_2_4              <dbl> 1, 2, 4, 3, 3, 2, 2, 1, 2, 4, 3, 3, 3, 3, 2~
$ STAI_pre_2_5              <dbl> 1, 2, 4, 1, 1, 2, 1, 1, 1, 3, 1, 2, 1, 2, 1~
$ STAI_pre_2_6              <dbl> 1, 3, 4, 1, 1, 2, 1, 1, 1, 3, 1, 1, 1, 2, 2~
$ STAI_pre_2_7              <dbl> 1, 1, 2, 2, 1, 1, 2, 1, 1, 1, 2, 1, 2, 3, 1~
$ STAI_pre_3_1              <dbl> 2, 3, 4, 3, 1, 2, 2, 1, 2, 4, 2, 2, 3, 2, 3~
$ STAI_pre_3_2              <dbl> 2, 3, 3, 3, 2, 2, 2, 1, 2, 3, 2, 2, 2, 3, 2~
$ STAI_pre_3_3              <dbl> 2, 3, 2, 2, 2, 2, 1, 1, 1, 3, 1, 1, 2, 1, 2~
$ STAI_pre_3_4              <dbl> 1, 2, 3, 1, 1, 1, 2, 1, 2, 3, 1, 1, 1, 1, 1~
$ STAI_pre_3_5              <dbl> 2, 3, 4, 3, 3, 2, 2, 1, 2, 4, 2, 2, 3, 2, 3~
$ STAI_pre_3_6              <dbl> 2, 3, 4, 3, 3, 2, 2, 1, 2, 4, 2, 2, 2, 2, 2~
$ STAI_post_1_1             <dbl> NA, 3, 3, 2, NA, 2, NA, NA, 2, 2, 2, 3, NA,~
$ STAI_post_1_2             <dbl> NA, 3, 3, 2, NA, 2, NA, NA, 2, 2, 2, 2, NA,~
$ STAI_post_1_3             <dbl> NA, 3, 2, 1, NA, 1, NA, NA, 1, 1, 2, 2, NA,~
$ STAI_post_1_4             <dbl> NA, 3, 2, 1, NA, 1, NA, NA, 1, 2, 1, 2, NA,~
$ STAI_post_1_5             <dbl> NA, 2, 2, 2, NA, 2, NA, NA, 2, 3, 3, 3, NA,~
$ STAI_post_1_6             <dbl> NA, 2, 1, 1, NA, 1, NA, NA, 1, 1, 3, 2, NA,~
$ STAI_post_1_7             <dbl> NA, 3, 1, 1, NA, 2, NA, NA, 1, 1, 1, 1, NA,~
$ STAI_post_2_1             <dbl> NA, 2, 3, 2, NA, 2, NA, NA, 2, 2, 3, 2, NA,~
```

```
$ STAI_post_2_2             <dbl> NA, 2, 1, 1, NA, 1, NA, NA, 1, 1, 1, 1, NA,~
$ STAI_post_2_3             <dbl> NA, 3, 3, 2, NA, 2, NA, NA, 2, 3, 3, 3, NA,~
$ STAI_post_2_4             <dbl> NA, 3, 3, 2, NA, 2, NA, NA, 2, 3, 3, 3, NA,~
$ STAI_post_2_5             <dbl> NA, 3, 1, 1, NA, 1, NA, NA, 1, 1, 1, 1, NA,~
$ STAI_post_2_6             <dbl> NA, 3, 1, 1, NA, 1, NA, NA, 1, 1, 2, 1, NA,~
$ STAI_post_2_7             <dbl> NA, 1, 1, 2, NA, 1, NA, NA, 1, 1, 2, 1, NA,~
$ STAI_post_3_1             <dbl> NA, 2, 3, 2, NA, 2, NA, NA, 2, 3, 3, 3, NA,~
$ STAI_post_3_2             <dbl> NA, 2, 3, 2, NA, 2, NA, NA, 2, 3, 4, 3, NA,~
$ STAI_post_3_3             <dbl> NA, 3, 1, 1, NA, 1, NA, NA, 1, 1, 1, 2, NA,~
$ STAI_post_3_4             <dbl> NA, 2, 1, 1, NA, 1, NA, NA, 1, 1, 2, 1, NA,~
$ STAI_post_3_5             <dbl> NA, 3, 3, 3, NA, 3, NA, NA, 2, 3, 3, 3, NA,~
$ STAI_post_3_6             <dbl> NA, 3, 3, 2, NA, 3, NA, NA, 2, 3, 3, 3, NA,~
$ moral_dilemma_dog         <dbl> 9, 9, 8, 8, 3, 9, 9, 9, 6, 6, 8, 7, 9, 9, 6~
$ moral_dilemma_wallet      <dbl> 9, 9, 7, 4, 9, 9, 5, 4, 9, 8, 7, 9, 9, 9, 7~
$ moral_dilemma_plane       <dbl> 8, 9, 8, 8, 9, 9, 7, 1, 3, 9, 9, 6, 9, 9, 4~
$ moral_dilemma_resume      <dbl> 7, 8, 5, 6, 5, 9, 3, 7, 9, 8, 5, 9, 8, 9, 7~
$ moral_dilemma_kitten      <dbl> 9, 9, 8, 9, 5, 8, 6, 9, 9, 9, 8, 9, 7, 9, 6~
$ moral_dilemma_trolley     <dbl> 5, 3, 5, 2, 4, 5, 3, 1, 1, 9, 2, 4, 5, 5, 3~
$ moral_dilemma_control     <dbl> 9, 2, 9, 8, 8, 6, 8, 7, 8, 6, 7, 8, 8, 3, 7~
$ presentation_experience   <dbl> NA, 2, 1, 2, NA, 2, NA, NA, 2, 2, 2, 2, NA,~
$ presentation_unpleasant   <dbl> NA, 63, 68, 32, NA, 63, NA, NA, 14, 54, 82,~
$ presentation_fun          <dbl> NA, 58, 26, 59, NA, 54, NA, NA, 78, 42, 7, ~
$ presentation_challenge    <dbl> NA, 58, 65, 80, NA, 50, NA, NA, 47, 64, 72,~
$ PBC_1                     <dbl> 3, NA, NA, NA, 3, NA, 3, 1, NA, NA, NA, NA,~
$ PBC_2                     <dbl> 3, NA, NA, NA, 3, NA, 3, 4, NA, NA, NA, NA,~
$ PBC_3                     <dbl> 5, NA, NA, NA, 3, NA, 3, 1, NA, NA, NA, NA,~
$ PBC_4                     <dbl> 5, NA, NA, NA, 3, NA, 5, 4, NA, NA, NA, NA,~
$ PBC_5                     <dbl> 5, NA, NA, NA, 2, NA, 5, 4, NA, NA, NA, NA,~
$ REI_1                     <dbl> 5, NA, NA, NA, 3, NA, 4, 4, NA, NA, NA, NA,~
$ REI_2                     <dbl> 4, NA, NA, NA, 3, NA, 4, 5, NA, NA, NA, NA,~
$ REI_3                     <dbl> 5, NA, NA, NA, 3, NA, 4, 5, NA, NA, NA, NA,~
$ REI_4                     <dbl> 4, NA, NA, NA, 3, NA, 4, 5, NA, NA, NA, NA,~
$ REI_5                     <dbl> 4, NA, NA, NA, 4, NA, 4, 5, NA, NA, NA, NA,~
$ REI_6                     <dbl> 5, NA, NA, NA, 3, NA, 4, 5, NA, NA, NA, NA,~
$ REI_7                     <dbl> 3, NA, NA, NA, 3, NA, 3, 5, NA, NA, NA, NA,~
$ REI_8                     <dbl> 4, NA, NA, NA, 3, NA, 4, 5, NA, NA, NA, NA,~
$ REI_9                     <dbl> 3, NA, NA, NA, 3, NA, 4, 3, NA, NA, NA, NA,~
$ REI_10                    <dbl> 4, NA, NA, NA, 3, NA, 4, 4, NA, NA, NA, NA,~
$ REI_11                    <dbl> 5, NA, NA, NA, 3, NA, 4, 5, NA, NA, NA, NA,~
$ REI_12                    <dbl> 5, NA, NA, NA, NA, NA, 3, 4, NA, NA, NA, NA~
$ REI_13                    <dbl> 3, NA, NA, NA, 3, NA, 4, 2, NA, NA, NA, NA,~
$ REI_14                    <dbl> 4, NA, NA, NA, 3, NA, 4, 2, NA, NA, NA, NA,~
$ REI_15                    <dbl> 4, NA, NA, NA, 3, NA, 3, 4, NA, NA, NA, NA,~
$ REI_16                    <dbl> 4, NA, NA, NA, 3, NA, 4, 4, NA, NA, NA, NA,~
$ REI_17                    <dbl> 3, NA, NA, NA, 3, NA, 3, 4, NA, NA, NA, NA,~
$ REI_18                    <dbl> 5, NA, NA, NA, 3, NA, 2, 5, NA, NA, NA, NA,~
```

```
$ REI_19                <dbl> 1, NA, NA, NA, 3, NA, 4, 3, NA, NA, NA, NA,~
$ REI_20                <dbl> 3, NA, NA, NA, 3, NA, 5, 5, NA, NA, NA, NA,~
$ REI_21                <dbl> 5, NA, NA, NA, 3, NA, 4, 5, NA, NA, NA, NA,~
$ REI_22                <dbl> 3, NA, NA, NA, 3, NA, 4, 3, NA, NA, NA, NA,~
$ REI_23                <dbl> 4, NA, NA, NA, 3, NA, 5, 5, NA, NA, NA, NA,~
$ REI_24                <dbl> 2, NA, NA, NA, 3, NA, 1, 5, NA, NA, NA, NA,~
$ REI_25                <dbl> 3, NA, NA, NA, 3, NA, 2, 3, NA, NA, NA, NA,~
$ REI_26                <dbl> 5, NA, NA, NA, 3, NA, 4, 5, NA, NA, NA, NA,~
$ REI_27                <dbl> 5, NA, NA, NA, 3, NA, 2, 1, NA, NA, NA, NA,~
$ REI_28                <dbl> 3, NA, NA, NA, 3, NA, 4, 3, NA, NA, NA, NA,~
$ REI_29                <dbl> 3, NA, NA, NA, 3, NA, 4, 4, NA, NA, NA, NA,~
$ REI_30                <dbl> 4, NA, NA, NA, 3, NA, 3, 4, NA, NA, NA, NA,~
$ REI_31                <dbl> 3, NA, NA, NA, 3, NA, 4, 5, NA, NA, NA, NA,~
$ REI_32                <dbl> 3, NA, NA, NA, 3, NA, 3, 1, NA, NA, NA, NA,~
$ REI_33                <dbl> 4, NA, NA, NA, 3, NA, 4, 4, NA, NA, NA, NA,~
$ REI_34                <dbl> 3, NA, NA, NA, 3, NA, 3, 4, NA, NA, NA, NA,~
$ REI_35                <dbl> 4, NA, NA, NA, 3, NA, 5, 5, NA, NA, NA, NA,~
$ REI_36                <dbl> 3, NA, NA, NA, 3, NA, 4, 4, NA, NA, NA, NA,~
$ REI_37                <dbl> 4, NA, NA, NA, 3, NA, 4, 3, NA, NA, NA, NA,~
$ REI_38                <dbl> 4, NA, NA, NA, 3, NA, 4, 4, NA, NA, NA, NA,~
$ REI_39                <dbl> 4, NA, NA, NA, 3, NA, 4, 3, NA, NA, NA, NA,~
$ REI_40                <dbl> 4, NA, NA, NA, 3, NA, 3, 5, NA, NA, NA, NA,~
$ MAIA_1_1              <dbl> 2, NA, NA, NA, 3, NA, 4, 3, NA, NA, NA, NA,~
$ MAIA_1_2              <dbl> 4, NA, NA, NA, 3, NA, 4, 4, NA, NA, NA, NA,~
$ MAIA_1_3              <dbl> 4, NA, NA, NA, 3, NA, 4, 4, NA, NA, NA, NA,~
$ MAIA_1_4              <dbl> 4, NA, NA, NA, 3, NA, 5, 4, NA, NA, NA, NA,~
$ MAIA_1_5              <dbl> 2, NA, NA, NA, 3, NA, 2, 5, NA, NA, NA, NA,~
$ MAIA_1_6              <dbl> 2, NA, NA, NA, 3, NA, 2, 2, NA, NA, NA, NA,~
$ MAIA_1_7              <dbl> 2, NA, NA, NA, 3, NA, 4, 3, NA, NA, NA, NA,~
$ MAIA_1_8              <dbl> 3, NA, NA, NA, 3, NA, 3, 5, NA, NA, NA, NA,~
$ MAIA_1_9              <dbl> 4, NA, NA, NA, 3, NA, 5, 6, NA, NA, NA, NA,~
$ MAIA_1_10             <dbl> 4, NA, NA, NA, 3, NA, 4, 2, NA, NA, NA, NA,~
$ MAIA_1_11             <dbl> 4, NA, NA, NA, 3, NA, 1, 3, NA, NA, NA, NA,~
$ MAIA_1_12             <dbl> 3, NA, NA, NA, 3, NA, 2, 3, NA, NA, NA, NA,~
$ MAIA_1_13             <dbl> 4, NA, NA, NA, 3, NA, 3, 5, NA, NA, NA, NA,~
$ MAIA_1_14             <dbl> 4, NA, NA, NA, 3, NA, 2, 5, NA, NA, NA, NA,~
$ MAIA_1_15             <dbl> 4, NA, NA, NA, 3, NA, 3, 4, NA, NA, NA, NA,~
$ MAIA_1_16             <dbl> 4, NA, NA, NA, 3, NA, 2, 5, NA, NA, NA, NA,~
$ MAIA_2_1              <dbl> 4, NA, NA, NA, 3, NA, 3, 5, NA, NA, NA, NA,~
$ MAIA_2_2              <dbl> 4, NA, NA, NA, 3, NA, 4, 5, NA, NA, NA, NA,~
$ MAIA_2_3              <dbl> 4, NA, NA, NA, 3, NA, 3, 2, NA, NA, NA, NA,~
$ MAIA_2_4              <dbl> 4, NA, NA, NA, 3, NA, 4, 4, NA, NA, NA, NA,~
$ MAIA_2_5              <dbl> 4, NA, NA, NA, 3, NA, 4, 4, NA, NA, NA, NA,~
$ MAIA_2_6              <dbl> 4, NA, NA, NA, 3, NA, 5, 4, NA, NA, NA, NA,~
$ MAIA_2_7              <dbl> 4, NA, NA, NA, 3, NA, 3, 2, NA, NA, NA, NA,~
$ MAIA_2_8              <dbl> 4, NA, NA, NA, 3, NA, 3, 3, NA, NA, NA, NA,~
```

```
$ MAIA_2_9                    <dbl> 4, NA, NA, NA, 3, NA, 4, 4, NA, NA, NA, NA,~
$ MAIA_2_10                   <dbl> 4, NA, NA, NA, 3, NA, 2, 4, NA, NA, NA, NA,~
$ MAIA_2_11                   <dbl> 4, NA, NA, NA, 3, NA, 3, 3, NA, NA, NA, NA,~
$ MAIA_2_12                   <dbl> 3, NA, NA, NA, 3, NA, 2, 0, NA, NA, NA, NA,~
$ MAIA_2_13                   <dbl> 3, NA, NA, NA, 3, NA, 3, 3, NA, NA, NA, NA,~
$ MAIA_2_14                   <dbl> 4, NA, NA, NA, 3, NA, 3, 4, NA, NA, NA, NA,~
$ MAIA_2_15                   <dbl> 4, NA, NA, NA, 3, NA, 4, 4, NA, NA, NA, NA,~
$ MAIA_2_16                   <dbl> 4, NA, NA, NA, 3, NA, 4, 4, NA, NA, NA, NA,~
$ STAI_pre                    <dbl> 32, 49, 65, 42, 33, 34, 32, 21, 31, 60, 34,~
$ STAI_post                   <dbl> NA, 51, 41, 32, NA, 33, NA, NA, 30, 38, 45,~
$ MAIA_noticing               <dbl> 14, NA, NA, NA, 12, NA, 17, 15, NA, NA, NA,~
$ MAIA_not_distracting        <dbl> 6, NA, NA, NA, 9, NA, 8, 10, NA, NA, NA, NA~
$ MAIA_not_worrying           <dbl> 11, NA, NA, NA, 9, NA, 12, 13, NA, NA, NA, ~
$ MAIA_attention_regulation   <dbl> 27, NA, NA, NA, 21, NA, 16, 30, NA, NA, NA,~
$ MAIA_emotional_awareness    <dbl> 20, NA, NA, NA, 15, NA, 20, 19, NA, NA, NA,~
$ MAIA_self_regulation        <dbl> 16, NA, NA, NA, 12, NA, 12, 13, NA, NA, NA,~
$ MAIA_body_listening         <dbl> 10, NA, NA, NA, 9, NA, 8, 6, NA, NA, NA, NA~
$ MAIA_trusting               <dbl> 12, NA, NA, NA, 9, NA, 11, 12, NA, NA, NA, ~
$ PBC                         <dbl> 21, NA, NA, NA, 14, NA, 19, 14, NA, NA, NA,~
$ REI_rational_ability        <dbl> 38, NA, NA, NA, 31, NA, 37, 38, NA, NA, NA,~
$ REI_rational_engagement     <dbl> 38, NA, NA, NA, NA, NA, 26, 41, NA, NA, NA,~
$ REI_experiental_ability     <dbl> 36, NA, NA, NA, 30, NA, 43, 43, NA, NA, NA,~
$ REI_experiental_engagement <dbl> 39, NA, NA, NA, 30, NA, 40, 38, NA, NA, NA,~
$ moral_judgment              <dbl> 8.000000, 7.000000, 7.142857, 6.428571, 6.1~
$ moral_judgment_disgust      <dbl> 8.666667, 9.000000, 8.000000, 8.333333, 5.6~
$ moral_judgment_non_disgust <dbl> 7.000000, 6.666667, 5.666667, 4.000000, 6.0~
$ presentation_evaluation     <dbl> NA, 3, 3, 4, NA, 4, NA, NA, 4, 3, 3, 3, NA,~
$ logbook                     <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
$ exclude                     <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
```

Use `glimpse()` to identify columns and column types.

What are the dimensions of the data set.

Exercise 1

You can combine the following tasks into a single statement.

Select columns

Select `gender`, `subject`, `age` and referring to the `STAI` questionnaire.

```
judgments |> select(gender, subject, age, matches("^STAI"))
```

```
# A tibble: 188 x 45
   gender subject   age STAI_pre_1_1 STAI_pre_1_2 STAI_pre_1_3 STAI_pre_1_4
   <chr>    <dbl> <dbl>        <dbl>        <dbl>        <dbl>        <dbl>
 1 female       2    24            2            1            2            2
 2 female       1    19            3            2            3            1
 3 female       3    19            4            3            3            3
 4 female       4    22            2            2            2            2
 5 female       7    22            1            1            1            1
 6 female       6    22            2            2            1            1
 7 female       5    18            2            2            1            1
 8 male         9    20            1            1            1            1
 9 female      16    21            2            2            1            1
10 female      13    19            4            2            3            3
# i 178 more rows
# i 38 more variables: STAI_pre_1_5 <dbl>, STAI_pre_1_6 <dbl>,
#   STAI_pre_1_7 <dbl>, STAI_pre_2_1 <dbl>, STAI_pre_2_2 <dbl>,
#   STAI_pre_2_3 <dbl>, STAI_pre_2_4 <dbl>, STAI_pre_2_5 <dbl>,
#   STAI_pre_2_6 <dbl>, STAI_pre_2_7 <dbl>, STAI_pre_3_1 <dbl>,
#   STAI_pre_3_2 <dbl>, STAI_pre_3_3 <dbl>, STAI_pre_3_4 <dbl>,
#   STAI_pre_3_5 <dbl>, STAI_pre_3_6 <dbl>, STAI_post_1_1 <dbl>, ...
```

Select all subjects with `STAI_pre` greater than 65 and `STAI_post` greater than 40

```
judgments |> filter(STAI_pre > 65 & STAI_post > 40)
```

```
# A tibble: 3 x 158
  start_date end_date finished condition subject gender   age mood_pre mood_post
  <chr>      <chr>       <dbl> <chr>       <dbl> <chr>  <dbl>    <dbl>     <dbl>
1 11/3/2014  11/3/20~        1 stress         22 female    18       13        37
2 11/3/2014  11/3/20~        1 stress         36 female    21       32        49
3 11/6/2014  11/6/20~        1 stress        159 female    18        9         0
# i 149 more variables: STAI_pre_1_1 <dbl>, STAI_pre_1_2 <dbl>,
#   STAI_pre_1_3 <dbl>, STAI_pre_1_4 <dbl>, STAI_pre_1_5 <dbl>,
#   STAI_pre_1_6 <dbl>, STAI_pre_1_7 <dbl>, STAI_pre_2_1 <dbl>,
#   STAI_pre_2_2 <dbl>, STAI_pre_2_3 <dbl>, STAI_pre_2_4 <dbl>,
#   STAI_pre_2_5 <dbl>, STAI_pre_2_6 <dbl>, STAI_pre_2_7 <dbl>,
#   STAI_pre_3_1 <dbl>, STAI_pre_3_2 <dbl>, STAI_pre_3_3 <dbl>,
#   STAI_pre_3_4 <dbl>, STAI_pre_3_5 <dbl>, STAI_pre_3_6 <dbl>, ...
```

Sort the observations by `STAI_pre` so that the subject with the highest score is on top.

```
judgments |> arrange(desc(STAI_pre))
```

```
# A tibble: 188 x 158
   start_date end_date  finished condition subject gender   age mood_pre
   <chr>      <chr>        <dbl> <chr>       <dbl> <chr>  <dbl>    <dbl>
 1 11/3/2014  11/3/2014        1 stress         22 female    18       13
 2 11/6/2014  11/6/2014        1 stress        159 female    18        9
 3 11/3/2014  11/3/2014        1 stress         36 female    21       32
 4 11/5/2014  11/5/2014        1 stress        109 female    18       22
 5 11/6/2014  11/6/2014        1 stress        127 female    18       15
 6 11/4/2014  11/4/2014        1 stress         75 male      18       42
 7 11/4/2014  11/4/2014        1 stress         75 male      18       42
 8 11/7/2014  11/7/2014        1 stress        169 female    18       42
 9 11/3/2014  11/3/2014        1 stress          3 female    19       22
10 11/4/2014  11/4/2014        1 stress         73 female    19       25
# i 178 more rows
# i 150 more variables: mood_post <dbl>, STAI_pre_1_1 <dbl>,
#   STAI_pre_1_2 <dbl>, STAI_pre_1_3 <dbl>, STAI_pre_1_4 <dbl>,
#   STAI_pre_1_5 <dbl>, STAI_pre_1_6 <dbl>, STAI_pre_1_7 <dbl>,
#   STAI_pre_2_1 <dbl>, STAI_pre_2_2 <dbl>, STAI_pre_2_3 <dbl>,
#   STAI_pre_2_4 <dbl>, STAI_pre_2_5 <dbl>, STAI_pre_2_6 <dbl>,
#   STAI_pre_2_7 <dbl>, STAI_pre_3_1 <dbl>, STAI_pre_3_2 <dbl>, ...
```

For better readability, move the subject, STAI_pre and STAI_post columns to the first positions.

```
judgments <- judgments |> relocate(c(STAI_pre, STAI_post), .before=start_date)
```

Check if the data set contains duplicated rows

> **Tip**
>
> Compare the number of rows in the raw input table with the number of rows in a de-duplicated table containing unique rows only.

```
stopifnot("Duplicates present in judgements." = nrow(judgments) - count(judgments) == 0)
```

Exercise 2

Create a new column called `STAI_pre_category`

It should contain "low", "normal" and "high" entries depending on the `STAI_pre` values:

- if `STAI_pre` is less than 25 assign "low",
- if `STAI_pre` is over 65 assign "high",
- for all other values assign "normal".

Arrange the table by `STAI_pre` in a descending manner.

```
judgments <- judgments |>
  mutate(
    STAI_pre_category = case_when(
      STAI_pre < 25 ~ "low",
      STAI_pre > 65 ~ "high",
      .default = "normal"
      ),
    .after = STAI_pre,
  ) |>
  arrange(desc(STAI_pre))
```

Display distinct values in `STAI_pre` and `STAI_pre_category`.

```
judgments |> distinct(STAI_pre, STAI_pre_category)
```

```
# A tibble: 46 x 2
   STAI_pre STAI_pre_category
      <dbl> <chr>
 1       70 high
 2       68 high
 3       67 high
 4       66 high
 5       65 normal
 6       62 normal
 7       61 normal
 8       60 normal
 9       59 normal
10       58 normal
# i 36 more rows
```

Normalize the values in the REI group

Divide all entries in the REI questionnaire by 5, the maximal value.

```r
judgments <- judgments |> mutate(across(matches("^REI"), ~ .x / 5))
judgments |>
  select(matches("^REI"))
```

```
# A tibble: 188 x 44
   REI_1 REI_2 REI_3 REI_4 REI_5 REI_6 REI_7 REI_8 REI_9 REI_10 REI_11 REI_12
   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  <dbl>  <dbl>  <dbl>
 1    NA    NA    NA    NA    NA    NA    NA    NA    NA     NA     NA     NA
 2    NA    NA    NA    NA    NA    NA    NA    NA    NA     NA     NA     NA
 3    NA    NA    NA    NA    NA    NA    NA    NA    NA     NA     NA     NA
 4    NA    NA    NA    NA    NA    NA    NA    NA    NA     NA     NA     NA
 5    NA    NA    NA    NA    NA    NA    NA    NA    NA     NA     NA     NA
 6    NA    NA    NA    NA    NA    NA    NA    NA    NA     NA     NA     NA
 7    NA    NA    NA    NA    NA    NA    NA    NA    NA     NA     NA     NA
 8    NA    NA    NA    NA    NA    NA    NA    NA    NA     NA     NA     NA
 9    NA    NA    NA    NA    NA    NA    NA    NA    NA     NA     NA     NA
10    NA    NA    NA    NA    NA    NA    NA    NA    NA     NA     NA     NA
# i 178 more rows
# i 32 more variables: REI_13 <dbl>, REI_14 <dbl>, REI_15 <dbl>, REI_16 <dbl>,
#   REI_17 <dbl>, REI_18 <dbl>, REI_19 <dbl>, REI_20 <dbl>, REI_21 <dbl>,
#   REI_22 <dbl>, REI_23 <dbl>, REI_24 <dbl>, REI_25 <dbl>, REI_26 <dbl>,
#   REI_27 <dbl>, REI_28 <dbl>, REI_29 <dbl>, REI_30 <dbl>, REI_31 <dbl>,
#   REI_32 <dbl>, REI_33 <dbl>, REI_34 <dbl>, REI_35 <dbl>, REI_36 <dbl>,
#   REI_37 <dbl>, REI_38 <dbl>, REI_39 <dbl>, REI_40 <dbl>, ...
```