Intro To R

1/12/23

Table of contents

1 Exercises

On this page you can find the exercises for the intro-to-R course.

2 dplyr exercises

```
library("tidyverse")
```

```
-- Attaching packages --------------------------------------- tidyverse 1.3.2 --
v ggplot2 3.4.0      v purrr   1.0.0
v tibble  3.1.8      v dplyr   1.0.10
v tidyr   1.2.1      v stringr 1.5.0
v readr   2.1.3      v forcats 0.5.2
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
```

```
# load data
adsl <- haven::read_sas("data/adsl.sas7bdat") %>%
  mutate_if(is_character, na_if, "")
```

2.1 Data wrangling with dplyr

Load the `adsl` data-frame and select the following variables:

- USUBJID
- ARM
- SEX
- AGE
- AGEU
- AGEGR1
- COUNTRY
- EOSSTT

```r
adsl %>%
  select(USUBJID, ARM, SEX, starts_with("AGE"), COUNTRY, EOSSTT)
```

```
# A tibble: 306 x 8
   USUBJID     ARM                  SEX     AGE AGEU  AGEGR1 COUNTRY EOSSTT
   <chr>       <chr>                <chr> <dbl> <chr>  <dbl> <chr>   <chr>
 1 01-701-1015 Placebo              F        63 YEARS      2 USA     COMPLETED
 2 01-701-1023 Placebo              M        64 YEARS      2 USA     DISCONTINU~
 3 01-701-1028 Xanomeline High Dose M        71 YEARS      3 USA     COMPLETED
 4 01-701-1033 Xanomeline Low Dose  M        74 YEARS      3 USA     DISCONTINU~
 5 01-701-1034 Xanomeline High Dose F        77 YEARS      3 USA     COMPLETED
 6 01-701-1047 Placebo              F        85 YEARS      3 USA     DISCONTINU~
 7 01-701-1057 Screen Failure       F        59 YEARS      2 USA     <NA>
 8 01-701-1097 Xanomeline Low Dose  M        68 YEARS      3 USA     COMPLETED
 9 01-701-1111 Xanomeline Low Dose  F        81 YEARS      3 USA     DISCONTINU~
10 01-701-1115 Xanomeline Low Dose  M        84 YEARS      3 USA     DISCONTINU~
# ... with 296 more rows
```

On the selected variables, include only patients in the placebo arm who are 66, 77, 88, or 99 years old.

```r
adsl %>%
  select(USUBJID, SEX, ARM, EOSSTT, starts_with("AGE")) %>%
  filter(ARM == "Placebo",
         AGE %in% c(6:9*11))
```

```
# A tibble: 5 x 7
  USUBJID     SEX   ARM     EOSSTT          AGE AGEU  AGEGR1
  <chr>       <chr> <chr>   <chr>         <dbl> <chr>  <dbl>
1 01-705-1059 F     Placebo DISCONTINUED     66 YEARS      3
2 01-708-1171 F     Placebo COMPLETED        77 YEARS      3
3 01-710-1368 F     Placebo COMPLETED        88 YEARS      3
4 01-714-1035 F     Placebo COMPLETED        88 YEARS      3
5 01-718-1139 M     Placebo COMPLETED        77 YEARS      3
```

Further include the variable TRTSDTM (datetime of first exposure to treatment) and sort the previous data-frame according to this variable from most recent to least recent first exposure.

```r
adsl %>%
  select(USUBJID, SEX, ARM, EOSSTT, starts_with("AGE"), TRTSDTM) %>%
  filter(ARM == "Placebo",
         AGE %in% c(6:9*11)) %>%
  arrange(desc(TRTSDTM))
```

```
# A tibble: 5 x 8
  USUBJID     SEX   ARM     EOSSTT          AGE AGEU  AGEGR1 TRTSDTM
  <chr>       <chr> <chr>   <chr>         <dbl> <chr>  <dbl> <dttm>
1 01-714-1035 F     Placebo COMPLETED        88 YEARS      3 2014-04-17 00:00:00
2 01-710-1368 F     Placebo COMPLETED        88 YEARS      3 2013-10-23 00:00:00
3 01-705-1059 F     Placebo DISCONTINUED     66 YEARS      3 2013-08-05 00:00:00
4 01-718-1139 M     Placebo COMPLETED        77 YEARS      3 2013-05-19 00:00:00
5 01-708-1171 F     Placebo COMPLETED        77 YEARS      3 2012-12-06 00:00:00
```

3 tidyr_exercises

```
library(tidyverse)
library(admiral)
library(admiral.test)
library(dplyr)
library(tidyr)

# load data
ex <- admiral_ex
dm <- admiral_dm
ds <- admiral_ds
suppds <- admiral_suppds
```

3.1 Pivoting with tidyr

Load the `ex` data-frame from `admiral_ex`and select the following variables:

- USUBJID
- EXTRT
- VISIT
- EXSTDTC

```
ex %>%
  select(USUBJID, EXTRT, VISIT, EXSTDTC)
```

```
# A tibble: 591 x 4
   USUBJID     EXTRT      VISIT    EXSTDTC
   <chr>       <chr>      <chr>    <chr>
 1 01-701-1015 PLACEBO    BASELINE 2014-01-02
 2 01-701-1015 PLACEBO    WEEK 2   2014-01-17
 3 01-701-1015 PLACEBO    WEEK 24  2014-06-19
```

```
 4 01-701-1023 PLACEBO    BASELINE 2012-08-05
 5 01-701-1023 PLACEBO    WEEK 2   2012-08-28
 6 01-701-1028 XANOMELINE BASELINE 2013-07-19
 7 01-701-1028 XANOMELINE WEEK 2   2013-08-02
 8 01-701-1028 XANOMELINE WEEK 24  2014-01-07
 9 01-701-1033 XANOMELINE BASELINE 2014-03-18
10 01-701-1034 XANOMELINE BASELINE 2014-07-01
# ... with 581 more rows
```

Using `pivot_wider()` create a table that would shaped this way

USUBJID	EXTRT	BASELINE	WEEK 2	WEEK 24
...	...	...	...	...

```
ex %>%
  select(USUBJID, EXTRT, VISIT, EXSTDTC) %>%
  pivot_wider(names_from = "VISIT", values_from = "EXSTDTC")
```

```
# A tibble: 254 x 5
   USUBJID     EXTRT      BASELINE   `WEEK 2`   `WEEK 24`
   <chr>       <chr>      <chr>      <chr>      <chr>
 1 01-701-1015 PLACEBO    2014-01-02 2014-01-17 2014-06-19
 2 01-701-1023 PLACEBO    2012-08-05 2012-08-28 <NA>
 3 01-701-1028 XANOMELINE 2013-07-19 2013-08-02 2014-01-07
 4 01-701-1033 XANOMELINE 2014-03-18 <NA>       <NA>
 5 01-701-1034 XANOMELINE 2014-07-01 2014-07-16 2014-12-18
 6 01-701-1047 PLACEBO    2013-02-12 2013-02-26 <NA>
 7 01-701-1097 XANOMELINE 2014-01-01 2014-01-16 2014-06-19
 8 01-701-1111 XANOMELINE 2012-09-07 <NA>       <NA>
 9 01-701-1115 XANOMELINE 2012-11-30 2012-12-14 <NA>
10 01-701-1118 PLACEBO    2014-03-12 2014-03-27 2014-08-28
# ... with 244 more rows
```

Load the `dm` data-frame from `admiral_dm`and select the following variables:

- USUBJID
- RACE
- SEX

```
  dm %>%
    select(USUBJID, RACE, SEX)
```

```
# A tibble: 306 x 3
   USUBJID     RACE  SEX
   <chr>       <chr> <chr>
 1 01-701-1015 WHITE F
 2 01-701-1023 WHITE M
 3 01-701-1028 WHITE M
 4 01-701-1033 WHITE M
 5 01-701-1034 WHITE F
 6 01-701-1047 WHITE F
 7 01-701-1057 WHITE F
 8 01-701-1097 WHITE M
 9 01-701-1111 WHITE F
10 01-701-1115 WHITE M
# ... with 296 more rows
```

Using `pivot_longer()` create a table that would shaped this way

USUBJID	VAR	VAL
1001	RACE	WHITE
1001	SEX	M

```
  dm %>%
    select(USUBJID, RACE, SEX) %>%
    pivot_longer(cols = c(RACE, SEX),
                 names_to = "VAR",
                 values_to = "VAL")
```

```
# A tibble: 612 x 3
   USUBJID     VAR   VAL
   <chr>       <chr> <chr>
 1 01-701-1015 RACE  WHITE
 2 01-701-1015 SEX   F
 3 01-701-1023 RACE  WHITE
 4 01-701-1023 SEX   M
 5 01-701-1028 RACE  WHITE
```

```
 6 01-701-1028 SEX   M
 7 01-701-1033 RACE  WHITE
 8 01-701-1033 SEX   M
 9 01-701-1034 RACE  WHITE
10 01-701-1034 SEX   F
# ... with 602 more rows
```

3.2 Joining using dplyr

Load the `ds` data-frame from `admiral_ds` and `suppds` data-frame from `admiral_suppds`. Prior to joining the two datasets together, we may need to do some cleaning of the data on `suppds`.

- Filter IDVAR for "DSSEQ"
- Mutate IDVARVAL from type character to type numeric.
- Select USUBJID IDVARVAL QNAM QLABEL QVAL

```
suppds <- suppds %>%
  filter(IDVAR == "DSSEQ") %>%
  mutate(IDVARVAL = as.numeric(IDVARVAL)) %>%
  select(USUBJID, IDVARVAL, QNAM, QLABEL, QVAL)

suppds
```

```
# A tibble: 3 x 5
  USUBJID     IDVARVAL QNAM    QLABEL                          QVAL
  <chr>          <dbl> <chr>   <chr>                           <chr>
1 01-703-1175        2 ENTCRIT PROTOCOL ENTRY CRITERIA NOT MET 16
2 01-705-1382        2 ENTCRIT PROTOCOL ENTRY CRITERIA NOT MET 25
3 01-708-1372        3 ENTCRIT PROTOCOL ENTRY CRITERIA NOT MET 16
```

Join the two tables together using `USUBJID` and `DSSEQ` as the key joining variables.

```
ds %>%
  left_join(suppds, by = c("USUBJID" = "USUBJID", "DSSEQ" = "IDVARVAL"))
```

```
# A tibble: 850 x 16
   STUDYID  DOMAIN USUBJID DSSEQ DSSPID DSTERM DSDECOD DSCAT VISIT~1 VISIT DSDTC
   <chr>    <chr>  <chr>   <dbl> <chr>  <chr>  <chr>   <chr>   <dbl> <chr> <chr>
 1 CDISCPI~ DS     01-701~     1 <NA>   RANDO~ RANDOM~ PROT~       3 BASE~ 2014~
 2 CDISCPI~ DS     01-701~     2 <NA>   PROTO~ COMPLE~ DISP~      13 WEEK~ 2014~
 3 CDISCPI~ DS     01-701~     3 <NA>   FINAL~ FINAL ~ OTHE~      13 WEEK~ 2014~
 4 CDISCPI~ DS     01-701~     1 <NA>   RANDO~ RANDOM~ PROT~       3 BASE~ 2012~
 5 CDISCPI~ DS     01-701~     2 24     ADVER~ ADVERS~ DISP~       5 WEEK~ 2012~
 6 CDISCPI~ DS     01-701~     3 <NA>   FINAL~ FINAL ~ OTHE~       5 WEEK~ 2012~
 7 CDISCPI~ DS     01-701~     4 <NA>   FINAL~ FINAL ~ OTHE~     201 RETR~ 2013~
 8 CDISCPI~ DS     01-701~     1 <NA>   RANDO~ RANDOM~ PROT~       3 BASE~ 2013~
 9 CDISCPI~ DS     01-701~     2 <NA>   PROTO~ COMPLE~ DISP~      13 WEEK~ 2014~
10 CDISCPI~ DS     01-701~     3 <NA>   FINAL~ FINAL ~ OTHE~      13 WEEK~ 2014~
# ... with 840 more rows, 5 more variables: DSSTDTC <chr>, DSSTDY <dbl>,
#   QNAM <chr>, QLABEL <chr>, QVAL <chr>, and abbreviated variable name
#   1: VISITNUM
```