

This is a replica of all the R scripts

Jianying

Last compiled 25 June, 2025

Getting sample ADaM data from PHUSE

```
adam_path <- "https://github.com/phuse-org/TestDataFactory/raw/main/Updated/TDF_ADaM/"
adsl <- haven::read_xpt(paste0(adam_path, "adsl.xpt"))

# Select a few variables
adsl <- adsl %>%
  dplyr::select(STUDYID, USUBJID, SUBJID, AGE, TRT01P, TRTSDT, TRTEDT, RACE, SEX, DISCONFL)
```

Take a quick peek

```
str(adsl)

## tibble [254 x 10] (S3: tbl_df/tbl/data.frame)
##  $ STUDYID : chr [1:254] "CDISCPIL0T01" "CDISCPIL0T01" "CDISCPIL0T01" "CDISCPIL0T01" ...
##  ..- attr(*, "label")= chr "Study Identifier"
##  $ USUBJID : chr [1:254] "01-701-1015" "01-701-1023" "01-701-1028" "01-701-1033" ...
##  ..- attr(*, "label")= chr "Unique Subject Identifier"
##  $ SUBJID  : chr [1:254] "1015" "1023" "1028" "1033" ...
##  ..- attr(*, "label")= chr "Subject Identifier for the Study"
##  $ AGE     : num [1:254] 63 64 71 74 77 85 68 81 84 52 ...
##  ..- attr(*, "label")= chr "Age"
##  $ TRT01P  : chr [1:254] "Placebo" "Placebo" "Xanomeline High Dose" "Xanomeline Low Dose" ...
##  ..- attr(*, "label")= chr "Planned Treatment for Period 01"
##  $ TRTSDT  : Date[1:254], format: "2014-01-02" "2012-08-05" ...
##  $ TRTEDT  : Date[1:254], format: "2014-07-02" "2012-09-01" ...
##  $ RACE    : chr [1:254] "WHITE" "WHITE" "WHITE" "WHITE" ...
##  ..- attr(*, "label")= chr "Race"
##  $ SEX     : chr [1:254] "F" "M" "M" "M" ...
##  ..- attr(*, "label")= chr "Sex"
##  $ DISCONFL: chr [1:254] "" "Y" "" "Y" ...
##  ..- attr(*, "label")= chr "Did the Subject Discontinue the Study?"

summary(adsl)
```

```
##      STUDYID          USUBJID          SUBJID          AGE
## Length:254          Length:254          Length:254          Min.   :51.00
## Class :character    Class :character    Class :character    1st Qu.:70.00
## Mode  :character    Mode  :character    Mode  :character    Median :77.00
##                                           Mean   :75.09
##                                           3rd Qu.:81.00
##                                           Max.   :89.00
##      TRT01P          TRTSDT          TRTEDT
## Length:254          Min.   :2012-07-09    Min.   :2012-08-28
## Class :character    1st Qu.:2013-01-26    1st Qu.:2013-05-12
## Mode  :character    Median :2013-06-13    Median :2013-09-27
##                                           Mean   :2013-06-17    Mean   :2013-10-10
##                                           3rd Qu.:2013-11-07    3rd Qu.:2014-03-15
##                                           Max.   :2014-09-02    Max.   :2015-03-05
##      RACE          SEX          DISCONFL
## Length:254          Length:254          Length:254
## Class :character    Class :character    Class :character
## Mode  :character    Mode  :character    Mode  :character
##
##
##
```

```
table(adsl$SEX)
```

```
##
##   F   M
## 143 111
```

```
adsl %>%
  count(SEX)
```

```
## # A tibble: 2 x 2
##   SEX      n
##   <chr> <int>
## 1 F      143
## 2 M      111
```

```
#for multiple variables
table(adsl$SEX, adsl$TRT01P, adsl$RACE)
```

```
## , , = AMERICAN INDIAN OR ALASKA NATIVE
##
##
##   Placebo Xanomeline High Dose Xanomeline Low Dose
##   F       0              0              0
##   M       0              1              0
##
## , , = BLACK OR AFRICAN AMERICAN
##
##
##   Placebo Xanomeline High Dose Xanomeline Low Dose
##   F       5              6              6
```

```
##      M      3      3      0
##
## , , = WHITE
##
##
##      Placebo Xanomeline High Dose Xanomeline Low Dose
##      F      48      34      44
##      M      30      40      34
```

```
adsl %>%
  count(SEX, TRT01P, RACE)
```

```
## # A tibble: 12 x 4
##   SEX   TRT01P      RACE      n
##   <chr> <chr>      <chr>   <int>
## 1 F     Placebo    BLACK OR AFRICAN AMERICAN     5
## 2 F     Placebo    WHITE                        48
## 3 F     Xanomeline High Dose BLACK OR AFRICAN AMERICAN     6
## 4 F     Xanomeline High Dose WHITE                        34
## 5 F     Xanomeline Low Dose  BLACK OR AFRICAN AMERICAN     6
## 6 F     Xanomeline Low Dose  WHITE                        44
## 7 M     Placebo    BLACK OR AFRICAN AMERICAN     3
## 8 M     Placebo    WHITE                        30
## 9 M     Xanomeline High Dose  AMERICAN INDIAN OR ALASKA NATIVE 1
## 10 M    Xanomeline High Dose  BLACK OR AFRICAN AMERICAN     3
## 11 M    Xanomeline High Dose  WHITE                        40
## 12 M    Xanomeline Low Dose   WHITE                        34
```

R processing that matches SAS

```
# to keep (only) SUBJID
adsl %>%
  select(SUBJID) %>%
  head() # keep only first 6 rows
```

```
## # A tibble: 6 x 1
##   SUBJID
##   <chr>
## 1 1015
## 2 1023
## 3 1028
## 4 1033
## 5 1034
## 6 1047
```

```
# to keep all variables except SUBJID
adsl %>%
  select(-SUBJID) %>%
  head()
```

```
## # A tibble: 6 x 9
##   STUDYID      USUBJID      AGE TRT01P TRTSDT      TRTEDT      RACE SEX DISCONFL
##   <chr>      <chr>      <dbl> <chr> <date>      <date>      <chr> <chr> <chr>
## 1 CDISCP~ 01-701~ 63 Place~ 2014-01-02 2014-07-02 WHITE F ""
## 2 CDISCP~ 01-701~ 64 Place~ 2012-08-05 2012-09-01 WHITE M "Y"
## 3 CDISCP~ 01-701~ 71 Xanom~ 2013-07-19 2014-01-14 WHITE M ""
## 4 CDISCP~ 01-701~ 74 Xanom~ 2014-03-18 2014-03-31 WHITE M "Y"
## 5 CDISCP~ 01-701~ 77 Xanom~ 2014-07-01 2014-12-30 WHITE F ""
## 6 CDISCP~ 01-701~ 85 Place~ 2013-02-12 2013-03-09 WHITE F "Y"
```

```
# select female
```

```
adsl %>%
  filter(SEX == "F") %>%
  head()
```

```
## # A tibble: 6 x 10
##   STUDYID USUBJID SUBJID      AGE TRT01P TRTSDT      TRTEDT      RACE SEX DISCONFL
##   <chr>   <chr>   <chr>      <dbl> <chr> <date>      <date>      <chr> <chr> <chr>
## 1 CDISCP~ 01-701~ 1015      63 Place~ 2014-01-02 2014-07-02 WHITE F ""
## 2 CDISCP~ 01-701~ 1034      77 Xanom~ 2014-07-01 2014-12-30 WHITE F ""
## 3 CDISCP~ 01-701~ 1047      85 Place~ 2013-02-12 2013-03-09 WHITE F "Y"
## 4 CDISCP~ 01-701~ 1111      81 Xanom~ 2012-09-07 2012-09-16 WHITE F "Y"
## 5 CDISCP~ 01-701~ 1133      81 Xanom~ 2012-10-28 2013-04-28 WHITE F ""
## 6 CDISCP~ 01-701~ 1146      75 Xanom~ 2013-05-20 2013-06-26 WHITE F "Y"
```

```
# multiple conditions:
```

```
adsl %>%
  filter(SEX == "F" & AGE > 70) %>%
  head()
```

```
## # A tibble: 6 x 10
##   STUDYID USUBJID SUBJID      AGE TRT01P TRTSDT      TRTEDT      RACE SEX DISCONFL
##   <chr>   <chr>   <chr>      <dbl> <chr> <date>      <date>      <chr> <chr> <chr>
## 1 CDISCP~ 01-701~ 1034      77 Xanom~ 2014-07-01 2014-12-30 WHITE F ""
## 2 CDISCP~ 01-701~ 1047      85 Place~ 2013-02-12 2013-03-09 WHITE F "Y"
## 3 CDISCP~ 01-701~ 1111      81 Xanom~ 2012-09-07 2012-09-16 WHITE F "Y"
## 4 CDISCP~ 01-701~ 1133      81 Xanom~ 2012-10-28 2013-04-28 WHITE F ""
## 5 CDISCP~ 01-701~ 1146      75 Xanom~ 2013-05-20 2013-06-26 WHITE F "Y"
## 6 CDISCP~ 01-701~ 1153      79 Place~ 2013-09-23 2014-03-16 WHITE F ""
```

```
# ascending
```

```
adsl %>%
  arrange(AGE)%>%
  head()
```

```
## # A tibble: 6 x 10
##   STUDYID USUBJID SUBJID      AGE TRT01P TRTSDT      TRTEDT      RACE SEX DISCONFL
##   <chr>   <chr>   <chr>      <dbl> <chr> <date>      <date>      <chr> <chr> <chr>
## 1 CDISCP~ 01-701~ 1341      51 Xanom~ 2013-01-05 2013-01-26 WHITE M "Y"
## 2 CDISCP~ 01-701~ 1118      52 Place~ 2014-03-12 2014-09-09 WHITE M ""
## 3 CDISCP~ 01-709~ 1007      54 Xanom~ 2012-07-31 2012-08-28 WHITE F "Y"
```

```
## 4 CDISCP~ 01-701~ 1180      56 Xanom~ 2013-02-12 2013-03-18 WHITE M      "Y"
## 5 CDISCP~ 01-701~ 1239      56 Xanom~ 2014-01-11 2014-07-10 WHITE M      ""
## 6 CDISCP~ 01-701~ 1287      56 Xanom~ 2014-01-25 2014-07-26 WHITE F      ""
```

```
# descending
adsl %>%
  arrange(-AGE)%>%
  head()
```

```
## # A tibble: 6 x 10
##   STUDYID USUBJID SUBJID   AGE TRT01P TRTSDT      TRTEDT      RACE SEX DISCONFL
##   <chr>   <chr>   <chr> <dbl> <chr> <date>   <date>   <chr> <chr> <chr>
## 1 CDISCP~ 01-710~ 1083    89 Place~ 2013-07-22 2013-08-01 WHITE F      "Y"
## 2 CDISCP~ 01-703~ 1295    88 Xanom~ 2013-11-21 2014-04-19 WHITE F      "Y"
## 3 CDISCP~ 01-710~ 1002    88 Xanom~ 2014-01-14 2014-01-18 WHITE M      "Y"
## 4 CDISCP~ 01-710~ 1368    88 Place~ 2013-10-23 2014-04-24 WHITE F      ""
## 5 CDISCP~ 01-714~ 1035    88 Place~ 2014-04-17 2014-10-16 WHITE F      ""
## 6 CDISCP~ 01-701~ 1387    87 Place~ 2014-03-12 2014-03-25 WHITE F      "Y"
```

```
## with case_when function
```

```
adsl %>%
  mutate(AGEGR1 = case_when(
    AGE > 50 ~"> 50 years old",
    AGE <= 50 ~"<= 50 years old",
  )) %>%
  head()
```

```
## # A tibble: 6 x 11
##   STUDYID USUBJID SUBJID   AGE TRT01P TRTSDT      TRTEDT      RACE SEX DISCONFL
##   <chr>   <chr>   <chr> <dbl> <chr> <date>   <date>   <chr> <chr> <chr>
## 1 CDISCP~ 01-701~ 1015    63 Place~ 2014-01-02 2014-07-02 WHITE F      ""
## 2 CDISCP~ 01-701~ 1023    64 Place~ 2012-08-05 2012-09-01 WHITE M      "Y"
## 3 CDISCP~ 01-701~ 1028    71 Xanom~ 2013-07-19 2014-01-14 WHITE M      ""
## 4 CDISCP~ 01-701~ 1033    74 Xanom~ 2014-03-18 2014-03-31 WHITE M      "Y"
## 5 CDISCP~ 01-701~ 1034    77 Xanom~ 2014-07-01 2014-12-30 WHITE F      ""
## 6 CDISCP~ 01-701~ 1047    85 Place~ 2013-02-12 2013-03-09 WHITE F      "Y"
## # i 1 more variable: AGEGR1 <chr>
```

Handling missing values

```
# Unlike SAS, space(s) != blank string in R,
"" == ""
```

```
## [1] FALSE
```

```
# This is true
"" == ""
```

```
## [1] TRUE
```

```
# is.na to check if it's missing
c(" ", "Y", NA_character_) %>% is.na()
```

```
## [1] FALSE FALSE TRUE
```

```
# NA is non-comparable, result in NA
c(" ", "Y", NA_character_) != "Y"
```

```
## [1] TRUE FALSE NA
```

```
# add one more row to the data with missing AGE and DISCONFL, select the first 3 rows
adsl_na_example <- adsl %>%
  add_row(SUBJID = "1", DISCONFL = NA_character_, .before = T) %>%
  slice(1:3) %>%
  select(SUBJID, DISCONFL)
# demo data
adsl_na_example %>% head()
```

```
## # A tibble: 3 x 2
##   SUBJID DISCONFL
##   <chr>   <chr>
## 1 1      <NA>
## 2 1015    ""
## 3 1023    "Y"
```

```
adsl_na_example %>% filter(DISCONFL != "Y")
```

```
## # A tibble: 1 x 2
##   SUBJID DISCONFL
##   <chr>   <chr>
## 1 1015    ""
```

```
# include is.na(DISCONFL) in the filter
adsl_na_example %>% filter(DISCONFL != "Y" | is.na(DISCONFL))
```

```
## # A tibble: 2 x 2
##   SUBJID DISCONFL
##   <chr>   <chr>
## 1 1      <NA>
## 2 1015    ""
```

```
## continue with sorting
```

```
# add one more row to the data with missing AGE and DISCONFL
# select the first 3 rows
adsl_na_sort <- adsl %>%
  add_row(SUBJID = "1", AGE = NA_integer_, .before = T) %>%
  slice(1:5) %>%
  select(SUBJID, AGE)
```

```
# ascending
adsl_na_sort %>%
  arrange(AGE)
```

```
## # A tibble: 5 x 2
##   SUBJID   AGE
##   <chr>  <dbl>
## 1 1015     63
## 2 1023     64
## 3 1028     71
## 4 1033     74
## 5 1       NA
```

```
# to be consistent with SAS - NA at the top when ascending
adsl_na_sort %>%
  arrange(!is.na(AGE), AGE)
```

```
## # A tibble: 5 x 2
##   SUBJID   AGE
##   <chr>  <dbl>
## 1 1       NA
## 2 1015     63
## 3 1023     64
## 4 1028     71
## 5 1033     74
```

```
SUBJID <- sample(adsl$SUBJID, 100, replace = FALSE)
WEIGHT <- sample(50:100, 100, replace = TRUE)
```

```
adsl_weight <- data.frame(SUBJID, WEIGHT)
```

```
# inner join
adsl %>%
  inner_join(adsl_weight, by = "SUBJID")
```

```
## # A tibble: 100 x 11
##   STUDYID      USUBJID   SUBJID   AGE TRT01P TRTSDT      TRTEDT      RACE SEX
##   <chr>        <chr>    <chr>  <dbl> <chr>  <date>    <date>    <chr> <chr>
## 1 CDISCPIL0T01 01-701-10~ 1015     63 Place~ 2014-01-02 2014-07-02 WHITE F
## 2 CDISCPIL0T01 01-701-10~ 1023     64 Place~ 2012-08-05 2012-09-01 WHITE M
## 3 CDISCPIL0T01 01-701-10~ 1028     71 Xanom~ 2013-07-19 2014-01-14 WHITE M
## 4 CDISCPIL0T01 01-701-10~ 1033     74 Xanom~ 2014-03-18 2014-03-31 WHITE M
## 5 CDISCPIL0T01 01-701-10~ 1034     77 Xanom~ 2014-07-01 2014-12-30 WHITE F
## 6 CDISCPIL0T01 01-701-10~ 1097     68 Xanom~ 2014-01-01 2014-07-09 WHITE M
## 7 CDISCPIL0T01 01-701-11~ 1111     81 Xanom~ 2012-09-07 2012-09-16 WHITE F
## 8 CDISCPIL0T01 01-701-11~ 1130     84 Place~ 2014-02-15 2014-08-16 WHITE M
## 9 CDISCPIL0T01 01-701-11~ 1181     79 Xanom~ 2013-12-05 2013-12-09 WHITE F
## 10 CDISCPIL0T01 01-701-12~ 1211     76 Xanom~ 2012-11-15 2013-01-12 WHITE F
## # i 90 more rows
## # i 2 more variables: DISCONFL <chr>, WEIGHT <int>
```

```
# outer join
adsl %>%
  full_join(adsl_weight, by = "SUBJID")
```

```
## # A tibble: 254 x 11
##   STUDYID      USUBJID   SUBJID   AGE TRT01P TRTSDT      TRTEDT    RACE SEX
##   <chr>        <chr>     <chr>  <dbl> <chr>  <date>    <date>    <chr> <chr>
## 1 CDISCPIL0T01 01-701-10~ 1015    63 Place~ 2014-01-02 2014-07-02 WHITE F
## 2 CDISCPIL0T01 01-701-10~ 1023    64 Place~ 2012-08-05 2012-09-01 WHITE M
## 3 CDISCPIL0T01 01-701-10~ 1028    71 Xanom~ 2013-07-19 2014-01-14 WHITE M
## 4 CDISCPIL0T01 01-701-10~ 1033    74 Xanom~ 2014-03-18 2014-03-31 WHITE M
## 5 CDISCPIL0T01 01-701-10~ 1034    77 Xanom~ 2014-07-01 2014-12-30 WHITE F
## 6 CDISCPIL0T01 01-701-10~ 1047    85 Place~ 2013-02-12 2013-03-09 WHITE F
## 7 CDISCPIL0T01 01-701-10~ 1097    68 Xanom~ 2014-01-01 2014-07-09 WHITE M
## 8 CDISCPIL0T01 01-701-11~ 1111    81 Xanom~ 2012-09-07 2012-09-16 WHITE F
## 9 CDISCPIL0T01 01-701-11~ 1115    84 Xanom~ 2012-11-30 2013-01-23 WHITE M
## 10 CDISCPIL0T01 01-701-11~ 1118    52 Place~ 2014-03-12 2014-09-09 WHITE M
## # i 244 more rows
## # i 2 more variables: DISCONFL <chr>, WEIGHT <int>
```

```
# left join
adsl %>%
  left_join(adsl_weight, by = "SUBJID")
```

```
## # A tibble: 254 x 11
##   STUDYID      USUBJID   SUBJID   AGE TRT01P TRTSDT      TRTEDT    RACE SEX
##   <chr>        <chr>     <chr>  <dbl> <chr>  <date>    <date>    <chr> <chr>
## 1 CDISCPIL0T01 01-701-10~ 1015    63 Place~ 2014-01-02 2014-07-02 WHITE F
## 2 CDISCPIL0T01 01-701-10~ 1023    64 Place~ 2012-08-05 2012-09-01 WHITE M
## 3 CDISCPIL0T01 01-701-10~ 1028    71 Xanom~ 2013-07-19 2014-01-14 WHITE M
## 4 CDISCPIL0T01 01-701-10~ 1033    74 Xanom~ 2014-03-18 2014-03-31 WHITE M
## 5 CDISCPIL0T01 01-701-10~ 1034    77 Xanom~ 2014-07-01 2014-12-30 WHITE F
## 6 CDISCPIL0T01 01-701-10~ 1047    85 Place~ 2013-02-12 2013-03-09 WHITE F
## 7 CDISCPIL0T01 01-701-10~ 1097    68 Xanom~ 2014-01-01 2014-07-09 WHITE M
## 8 CDISCPIL0T01 01-701-11~ 1111    81 Xanom~ 2012-09-07 2012-09-16 WHITE F
## 9 CDISCPIL0T01 01-701-11~ 1115    84 Xanom~ 2012-11-30 2013-01-23 WHITE M
## 10 CDISCPIL0T01 01-701-11~ 1118    52 Place~ 2014-03-12 2014-09-09 WHITE M
## # i 244 more rows
## # i 2 more variables: DISCONFL <chr>, WEIGHT <int>
```

```
# first 5 records of adsl
adsl_1 <- adsl[c(1:5),]

# the next 5 records of adsl
adsl_2 <- adsl[c(6:10),]

# tidyverse method - bind_rows()
adsl_stacked <- adsl_1 %>%
  bind_rows(adsl_2)
```


Clinical trial data

```
library(help = "datasets")
data("ToothGrowth")

library(haven)

ads1 <- read_xpt("https://github.com/phuse-org/TestDataFactory/raw/main/Updated/TDF_ADaM/ads1.xpt")
ads1 %>% dim()

## [1] 254 49
```

Work on ADaM data and build table with Tplyr

```
ads1 <- read_xpt("https://github.com/phuse-org/TestDataFactory/raw/main/Updated/TDF_ADaM/ads1.xpt")

## string formatting with Tplyr

library(Tplyr)

options(

  # Categorical variable defaults
  tplyr.count_layer_default_formats =
    list(n_counts = f_str("xxx [xx.xx%]", n, pct)),

  # Continuous variable defaults
  tplyr.desc_layer_default_formats =
    list('N' = f_str('xx', n),
         'Mean [SD]' = f_str('xx.xx [xx.xxx]', mean, sd),
         'Median' = f_str('xx.x', median),
         'Min, Max' = f_str('xx, xx', min, max))
)

library(dplyr)

# Initiate Tplyr, specify treatment variable, optional where condition

my_table <- tplyr_table(ads1, TRT01P, where = SAFFL == "Y") %>%

  # Add a total group column
  add_total_group() %>%

  # Add individual variables here
  add_layer(group_desc(AGE, b = "Age (years)")) %>%
  add_layer(group_count(AGEGR1, b = "Age Group 1 (years)")) %>%
  add_layer(group_count(SEX, b = "Gender")) %>%
  add_layer(group_count(ETHNIC, b = "Ethnicity")) %>%
```

```

add_layer(group_desc(BMIBL, b = "Baseline Body Mass Index (kg/m2)")) %>%

# Build
build()

head(my_table, n = 9)

## # A tibble: 9 x 9
##   row_label1      row_label2 var1_Placebo   `var1_Xanomeline High Dose`
##   <chr>          <chr>      <chr>          <chr>
## 1 Age (years)    N          "86"          "84"
## 2 Age (years)    Mean [SD]  "75.21 [ 8.590]" "74.38 [ 7.886]"
## 3 Age (years)    Median     "76.0"        "76.0"
## 4 Age (years)    Min, Max   "52, 89"      "56, 88"
## 5 Age Group 1 (years) <65      " 14 [16.28%]" " 11 [13.10%]"
## 6 Age Group 1 (years) >80        " 30 [34.88%]" " 18 [21.43%]"
## 7 Age Group 1 (years) 65-80      " 42 [48.84%]" " 55 [65.48%]"
## 8 Gender         F          " 53 [61.63%]" " 40 [47.62%]"
## 9 Gender         M          " 33 [38.37%]" " 44 [52.38%]"
## # i 5 more variables: `var1_Xanomeline Low Dose` <chr>, var1_Total <chr>,
## #   ord_layer_index <int>, ord_layer_1 <int>, ord_layer_2 <dbl>

my_table <- my_table %>%

# remove repeating labels
apply_row_masks(., row_breaks = TRUE) %>%

# specify order of relevant variables
select(row_label1,
       row_label2,
       `var1_Xanomeline High Dose`,
       `var1_Xanomeline Low Dose`,
       var1_Placebo,
       var1_Total)

library(flextable)

## a basic flextable
my_flextable <- my_table %>%
# start flextable
flextable() %>%
autofit()

# my_flextable

## a nicer flextable
my_flextable <- my_table %>%

# start flextable
flextable() %>%
autofit() %>%

# add some padding between rows

```

```
padding(padding = 0.5) %>%

# adjust width of first two columns
width(j = 1:2, width = 0.5) %>%

# align treatment columns to center
align(part = "all", align = "center", j = 3:6) %>%

# column header labels
set_header_labels(., values = list(
  row_label1 = 'Variable',
  row_label2 = ' ',
  `var1_Xanomeline High Dose` = 'Xanomeline \nHigh Dose',
  `var1_Xanomeline Low Dose` = 'Xanomeline \nLow Dose',
  var1_Placebo = 'Placebo',
  var1_Total = 'Total')) %>%

# header + footers
add_header_lines(values = "Table: Demographics (Safety Analysis Set)") %>%
add_footer_lines(values = "This was produced in R!") %>%

# font size, font name
fontsize(part = "all", size = 8)
# font()
# font(part = "all", fontname = "Times")
# This errors, perhaps version issue. -- SZ

ft <- my_flextable
ft <- autofit(ft)
# ft

ft <- set_table_properties(ft, width = .5, layout = "autofit")
ft
```

Table: Demographics (Safety Analysis Set)

Variable		Xanomeline High Dose	Xanomeline Low Dose	Placebo	Total
Age (years)	N	84	84	86	254
	Mean [SD]	74.38 [7.886]	75.67 [8.286]	75.21 [8.590]	75.09 [8.246]
	Median	76.0	77.5	76.0	77.0
	Min, Max	56, 88	51, 88	52, 89	51, 89
Age Group 1 (years)	<65	11 [13.10%]	8 [9.52%]	14 [16.28%]	33 [12.99%]
	>80	18 [21.43%]	29 [34.52%]	30 [34.88%]	77 [30.31%]
	65-80	55 [65.48%]	47 [55.95%]	42 [48.84%]	144 [56.69%]
Gender	F	40 [47.62%]	50 [59.52%]	53 [61.63%]	143 [56.30%]
	M	44 [52.38%]	34 [40.48%]	33 [38.37%]	111 [43.70%]

Table: Demographics (Safety Analysis Set)

Variable		Xanomeline High Dose	Xanomeline Low Dose	Placebo	Total
Ethnicity	HISPANIC OR LATINO	3 [3.57%]	6 [7.14%]	3 [3.49%]	12 [4.72%]
	NOT HISPANIC OR LATINO	81 [96.43%]	78 [92.86%]	83 [96.51%]	242 [95.28%]
Baseline Body Mass Index (kg/m2) N		84	84	86	254
	Mean [SD]	25.35 [4.158]	25.06 [4.271]	23.64 [3.672]	24.67 [4.092]
	Median	24.8	24.3	23.4	24.2
	Min, Max	14, 34	18, 40	15, 33	14, 40

This was produced in R!

```
library(officer)

# send the flextable to .docx output
read_docx() %>%
  body_add_flextable(ft) %>%
  print(target = "my_document.docx")
```

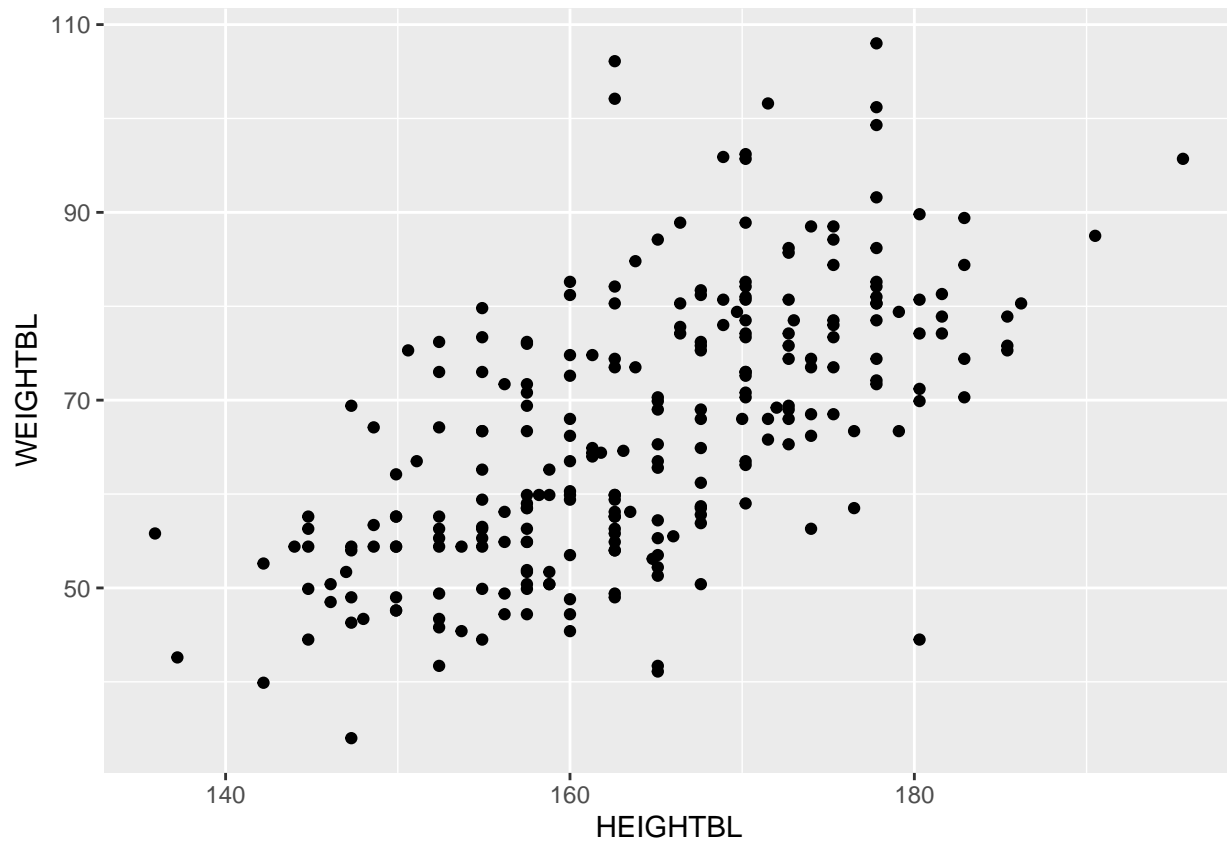
Produce figures

```
# Packages
library(haven)
library(ggplot2)

# Data
adam_path <- "https://github.com/phuse-org/TestDataFactory/raw/main/Updated/TDF_ADaM/"
adsl <- as.data.frame(haven::read_xpt(paste0(adam_path, "adsl.xpt")))

# basic scatter plot of weight vs height
my_scatter_plot <- ggplot(adsl, aes(x = HEIGHTBL, y = WEIGHTBL)) +
  geom_point()
my_scatter_plot
```

```
## Warning: Removed 1 row containing missing values or values outside the scale range
## (`geom_point()`).
```



```
# save plot as a png file
ggsave(plot = my_scatter_plot,
        filename = "my_plot_output_1.png",
        width = 7,
        height = 5,
        units = "in")
```

```
## Warning: Removed 1 row containing missing values or values outside the scale range
## (`geom_point()`).
```

```
png(filename = "my_plot_output_2.png",
     width = 7,
     height = 5,
     units = "in",
     res = 300)
```

```
print(my_scatter_plot)
```

```
## Warning: Removed 1 row containing missing values or values outside the scale range
## (`geom_point()`).
```

```
dev.off()
```

```
## pdf
## 2
```

phuse data for plot

```
# Packages
library(haven)
library(swimplot)
library(dplyr)

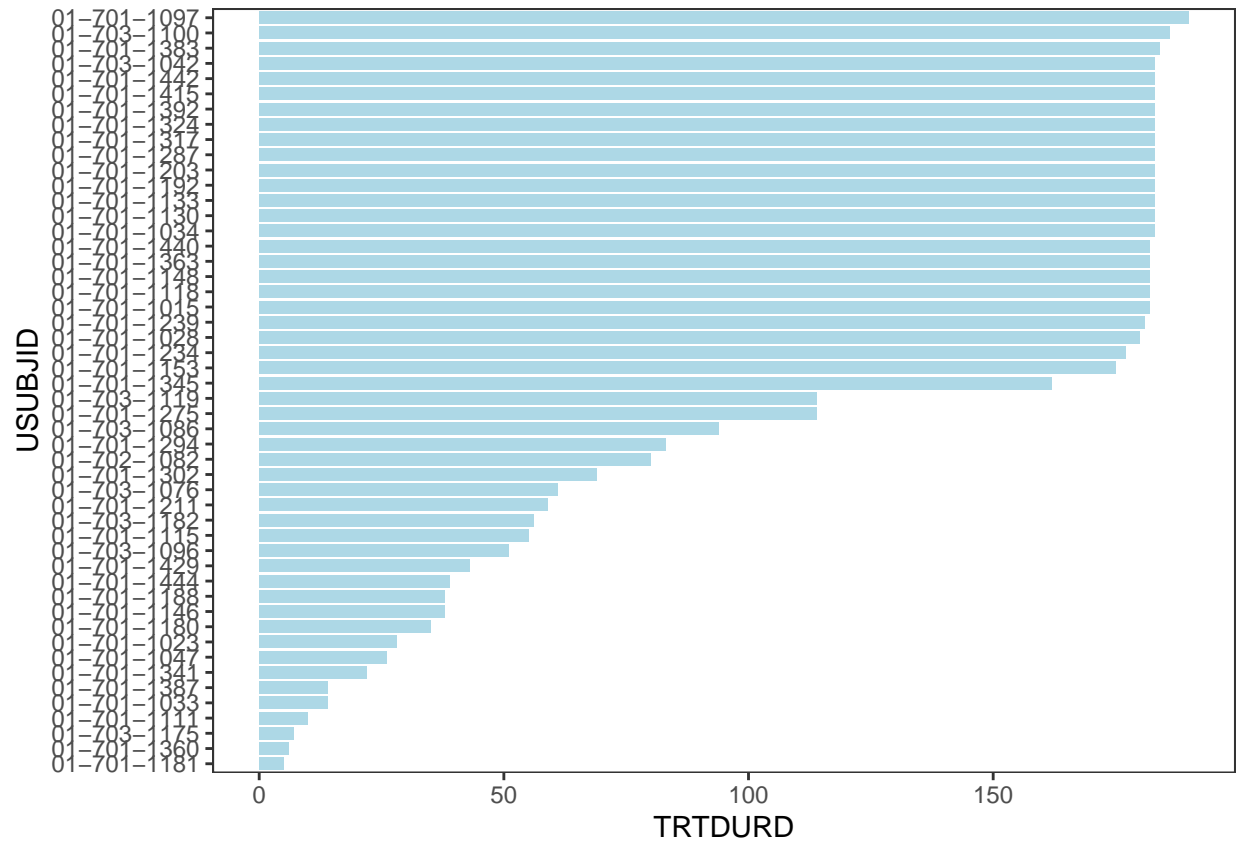
# Data
adam_path <- "https://github.com/phuse-org/TestDataFactory/raw/main/Updated/TDF_ADaM/"
adsl <- haven::read_xpt(paste0(adam_path, "adsl.xpt"))
adae <- haven::read_xpt(paste0(adam_path, "adae.xpt"))
adtte <- haven::read_xpt(paste0(adam_path, "adtte.xpt"))

adsl_new <- adsl %>%
  select(USUBJID, ARM, TRTDURD, SEX) %>%
  slice(1:50)
adae_new <- adae %>%
  select(USUBJID, AEDECOD, AESEV, AEREL, ASTDY) %>%
  filter(USUBJID %in% adsl_new$USUBJID & ASTDY >= 0)
adtte_new <- adtte %>%
  select(USUBJID, EVNTDESC, AVAL) %>%
  filter(USUBJID %in% adsl_new$USUBJID & EVNTDESC != "Study Completion Date")
random_duration_of_events <- sample(1:25, nrow(adtte_new), replace = T)
adtte_new <- adtte_new %>%
  bind_cols(random_duration_of_events) %>%
  mutate(Resp_end = AVAL + random_duration_of_events )

## New names:
## * `` -> `...4`

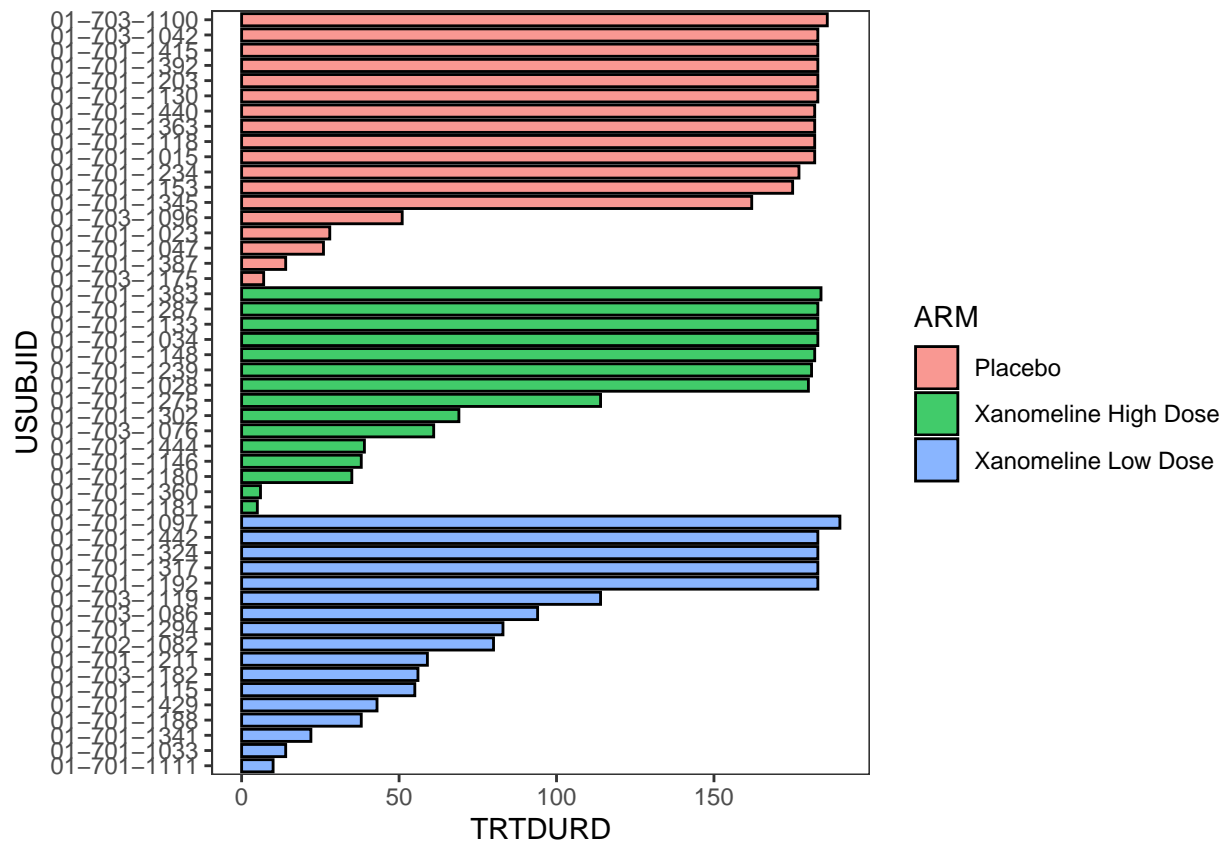
adsl_new <- as.data.frame(adsl_new)
adae_new <- as.data.frame(adae_new)
adtte_new <- as.data.frame(adtte_new)

swimmer_plot(df=adsl_new,
             id='USUBJID',
             end='TRTDURD',
             fill='lightblue',
             width=.85)
```



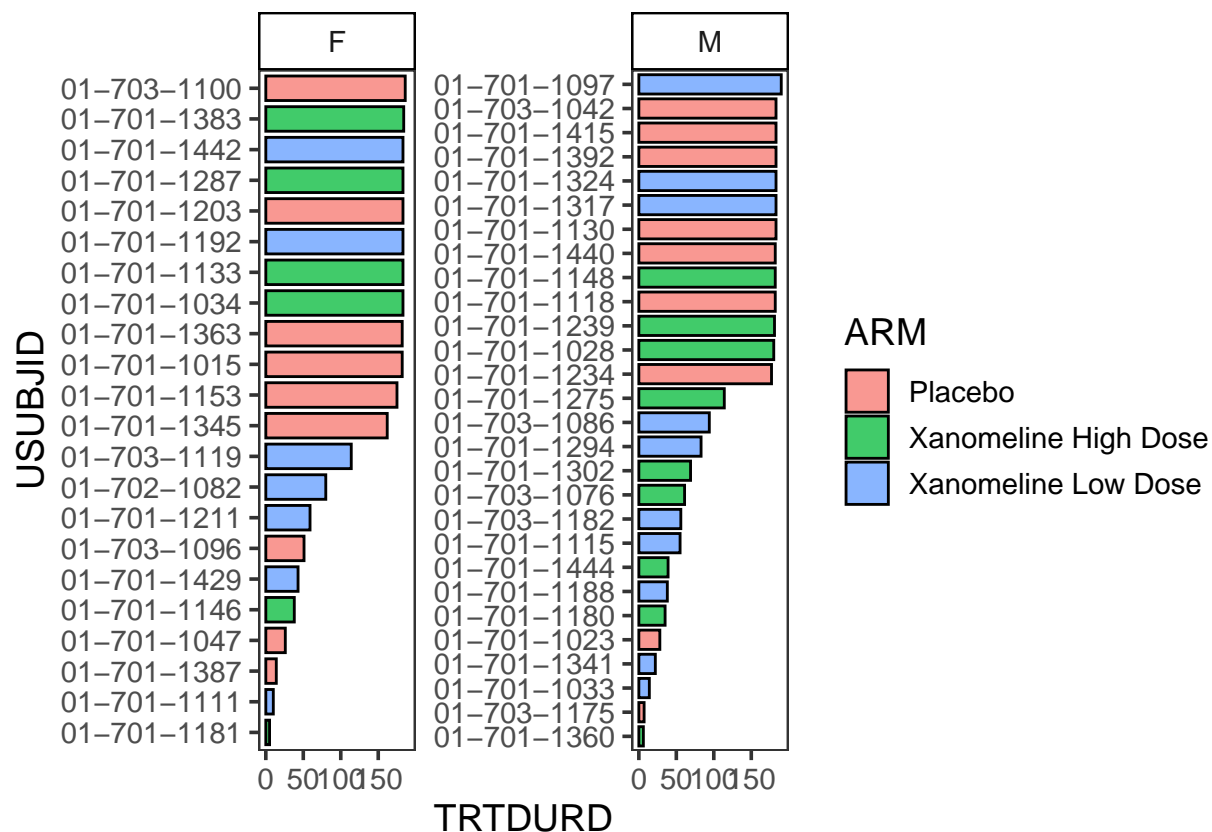
```
arm_plot <- swimmer_plot(df=adsl_new,
                        id='USUBJID',
                        end='TRTDURD',
                        name_fill='ARM',
                        id_order='ARM',
                        col="black",
                        alpha=0.75,
                        width=.8)

arm_plot
```

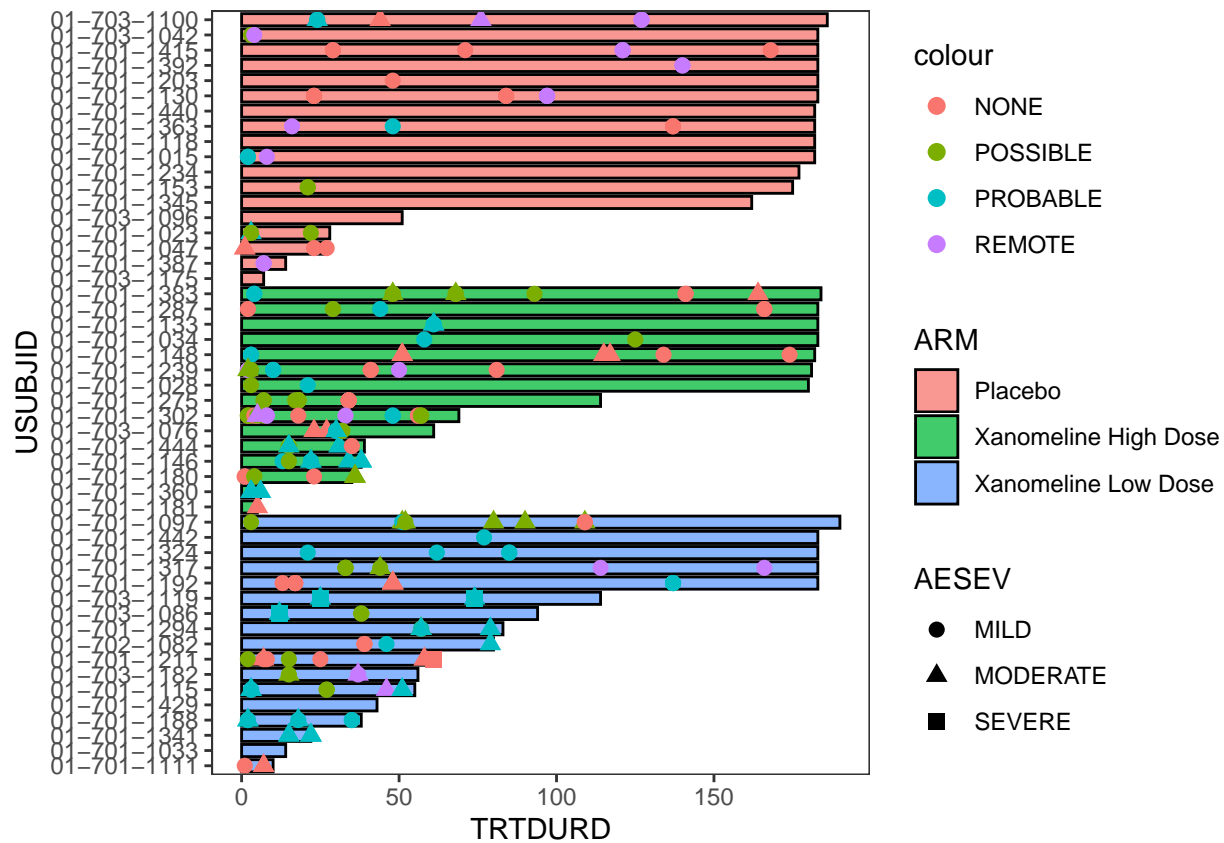


```
swim_plot_stratify <- swimmer_plot(df=adsl_new,
                                   id='USUBJID',
                                   end='TRTDURD',
                                   name_fill='ARM',
                                   col="black",
                                   alpha=0.75,
                                   width=.8,
                                   base_size=14,
                                   stratify= c('SEX'))

swim_plot_stratify
```

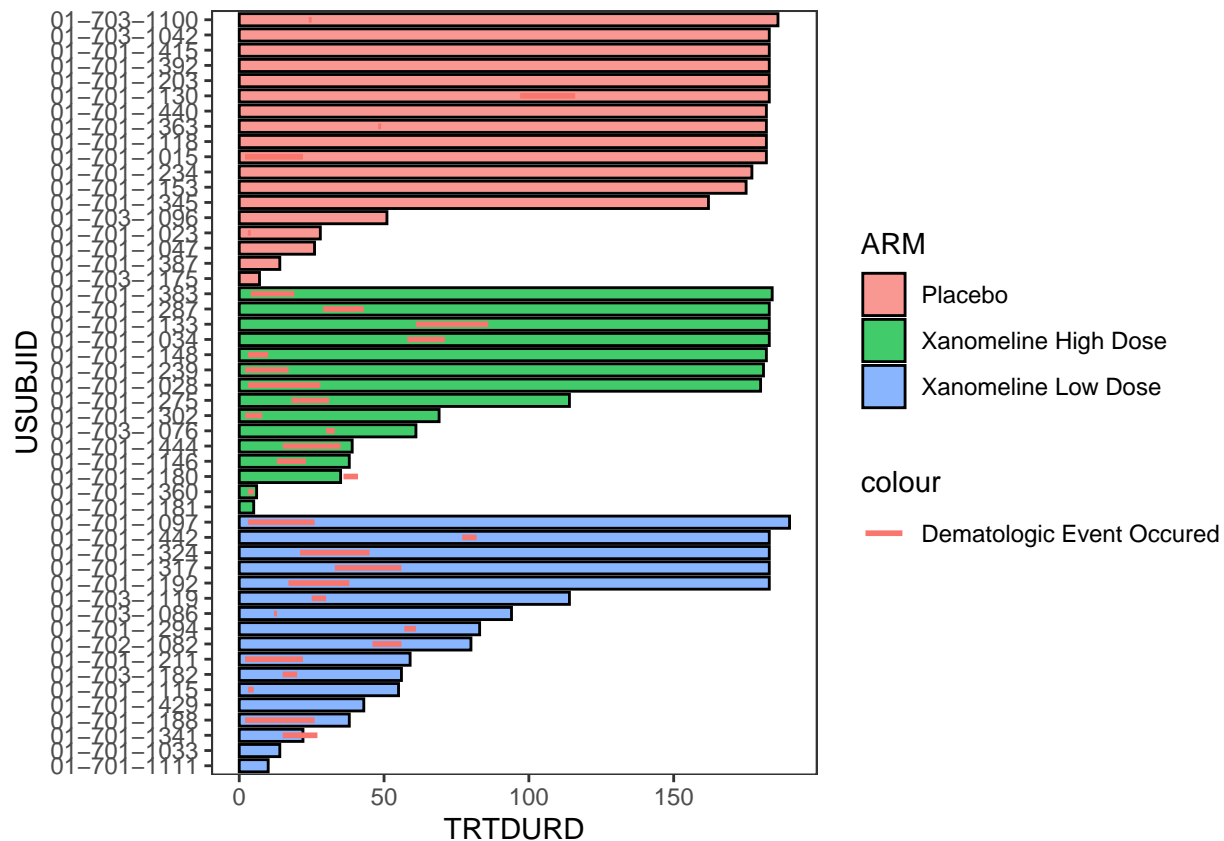



```
AE_plot <- arm_plot +
  swimmer_points(df_points=adae_new,
    id='USUBJID',
    time='ASTDY',
    name_shape='AESEV',
    size=2.5,
    fill='white',
    name_col='AEREL')
AE_plot
```



```
Response_plot <- arm_plot +
  swimmer_lines(df_lines=adtte_new,
    id='USUBJID',
    start='AVAL',
    end='Resp_end',
    name_col='EVNTDESC',
    size=1)
```

Response_plot

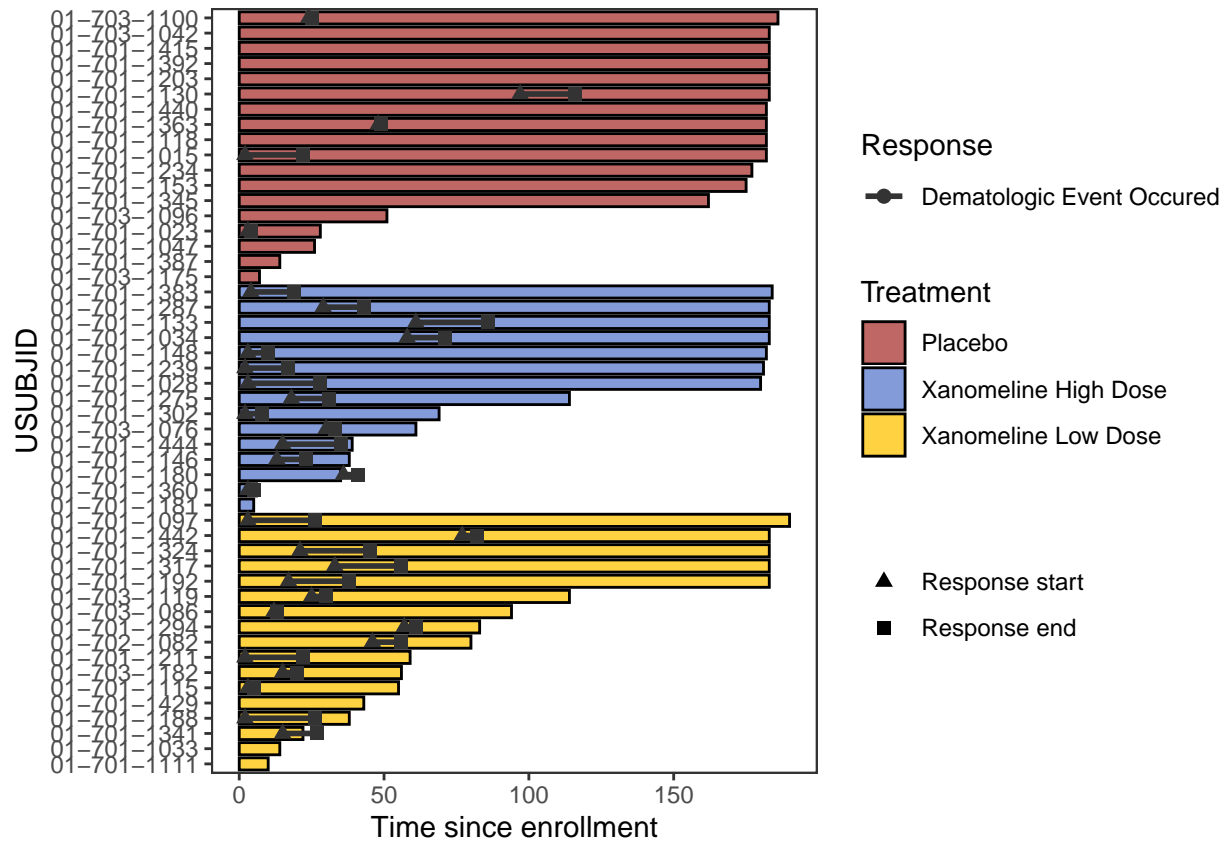


```
Response_plot_with_points <- Response_plot +
  swimmer_points_from_lines(df_lines=adtte_new,
    id='USUBJID',
    start='AVAL',
    end='Resp_end',
    name_col='EVNTDESC',
    size=2) +
  scale_fill_manual(name="Treatment",
    values=c("Placebo" = "#A9342F",
      "Xanomeline High Dose" = "#5B7ACE",
      "Xanomeline Low Dose" = "#FFC300')) +
  scale_color_manual(name="Response",
    values=c("grey20")) +
  scale_shape_manual(name='',
    values=c(17,15),
    breaks=c('AVAL', 'Resp_end'),
    labels=c('Response start', 'Response end')) +
  guides(fill = guide_legend(override.aes = list(shape = NA))) +
  scale_y_continuous(name = "Time since enrollment")
```

```
## Warning: `gather()` was deprecated in tidyr 1.2.0.
## i Please use `gather()` instead.
## i The deprecated feature was likely used in the swmplot package.
## Please report the issue to the authors.
## This warning is displayed once every 8 hours.
```

```
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```
Response_plot_with_points
```



```
# Packages
library(gridExtra)
```

```
##
## Attaching package: 'gridExtra'
##
## The following object is masked from 'package:dplyr':
##
##   combine
```

```
library(grid)
```

```
# Data
wp <- data.frame(subjidn = 1:30,
                  trtp = sample(c('Drug', 'Placebo'), replace = T, 30),
                  aval = runif(30, min = -40, max = 40))

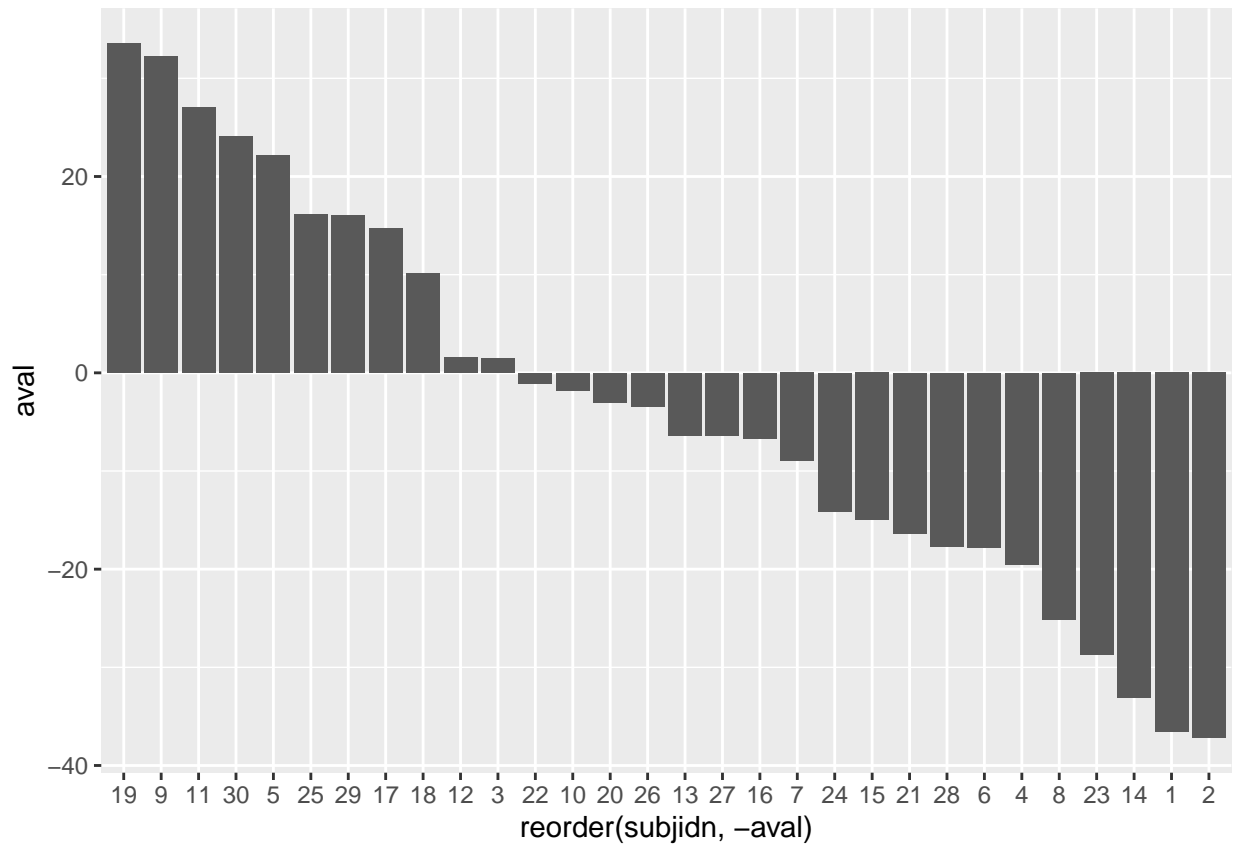
wp %>% head()
```

```
##   subjidn   trtp   aval
```

```
## 1      1      Drug -36.591249
## 2      2 Placebo -37.194116
## 3      3 Placebo  1.434185
## 4      4      Drug -19.553365
## 5      5      Drug  22.130284
## 6      6 Placebo -17.778419
```

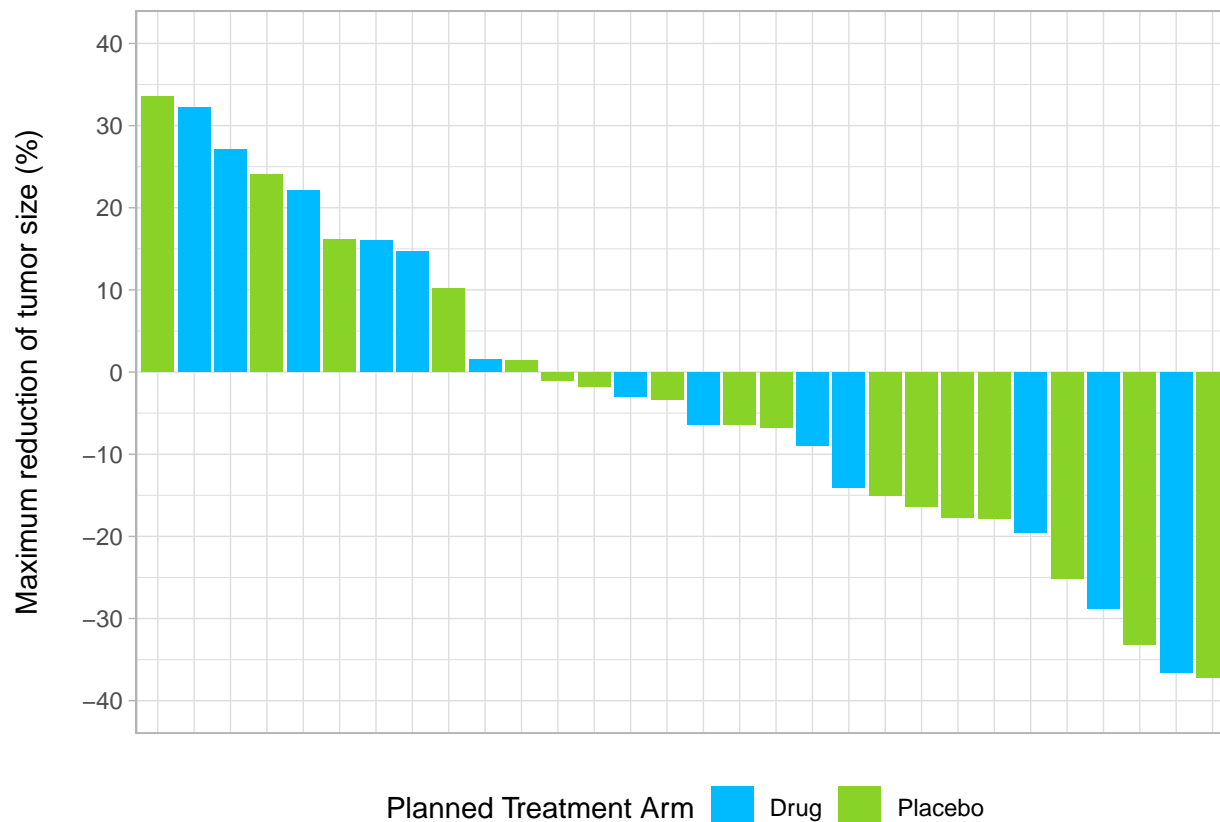
```
basic_waterfall <- ggplot(wp, aes(y = aval, x = reorder(subjdn, -aval))) +
  geom_bar(stat = "identity")

basic_waterfall
```



```
custom_waterfall <- ggplot(wp, aes(y = aval, x = reorder(subjdn, -aval), fill = trtp)) +
  geom_bar(stat = "identity") +
  scale_fill_manual("Planned Treatment Arm", values = c('#00bbff', '#89d329')) +
  scale_y_continuous(limits = c(-40, 40), breaks = seq(-40, 40, by = 10)) +
  ylab("Maximum reduction of tumor size (%)\\n") +
  theme_light() +
  theme(axis.title.x = element_blank(),
        axis.line.x = element_blank(),
        axis.text.x = element_blank(),
        axis.ticks.x = element_blank(),
        legend.position = "bottom")

custom_waterfall
```



on a simulated data

```
set.seed(100)

n <- 120 # size/records of simulated data

dat_all <-
  data.frame(SUBJID = 120010100:(120010100+n-1),
    AVAL = c(rnorm(round(0.8*n), 0, 20), rep(0, round(0.2*n))) %>% sample(),
    OVERALLRESP = c("PR", "SD", "PD") %>% sample(size=n, replace=TRUE),
    AMEDGRPN = seq(10, 50, 10) %>% sample(size=n, replace=TRUE),
    DOR = rpois(n, 40),
    AVISIT = c("C1", "C2", "C3", "EOT") %>% sample(size=n, replace=TRUE),
    ATMLOSS_L = c("E", "L", "Data unavailable") %>% sample(size=n, replace=TRUE),
    ATMLOSS_P = c("E", "L", "Data unavailable") %>% sample(size=n, replace=TRUE),
    ATMLOSS_R = c("E", "L", "Data unavailable") %>% sample(size=n, replace=TRUE),
    IDFOOT = "Bayer: /study/path/.../f_waterfall.sas" 30NOV2021 16:06"
  ) %>%
  mutate_at(vars("OVERALLRESP", "DOR", "IDFOOT"), as.character)

waterfall.plot <- dat_all %>% ggplot(aes(reorder(SUBJID, -AVAL), AVAL, fill =OVERALLRESP)) +
  geom_bar(stat="identity") +
  labs(x = "Subject",
```

```

    y = "Best % Change from baseline \n in sum of diameters \n (target lesions)\n" +
scale_x_discrete(labels = function(x) stringr::str_wrap(gsub("([0-9])([0-9])", "\\1 \\2 ", x),
                                width = 5) %>%
                                stringr::str_replace_all(" ", "")) +
scale_fill_manual("Best Response",
                  breaks = c("PR", "SD", "PD"),
                  values=c("PR"="#89d329",
                           "SD"="#756bb1",
                           "PD"="#00bbff")) +
geom_text(aes(label = if_else(AVAL == 0, OVERALLRESP, ""), fontface="bold"),
          vjust = -1,
          size=3,
          color="black") + theme_bw() +
theme(
  axis.text = element_text(size=12, color="black", face = "bold"),
  axis.title.y = element_text(size=12, face="bold"),
  axis.title.x = element_blank(),
  legend.background = element_rect(color = "steelblue", linetype = "solid"),
  legend.justification = c(1, 1),
  legend.position = c(1, 1),
  legend.direction = "horizontal",
  legend.text = element_text(size=8, color = "black", face = "bold"),
  legend.title = element_text(size=8, color = "black", face = "bold"),
  plot.caption = element_text(hjust = 0, size = 10, color = "blue"),
  plot.caption.position = "plot"
)

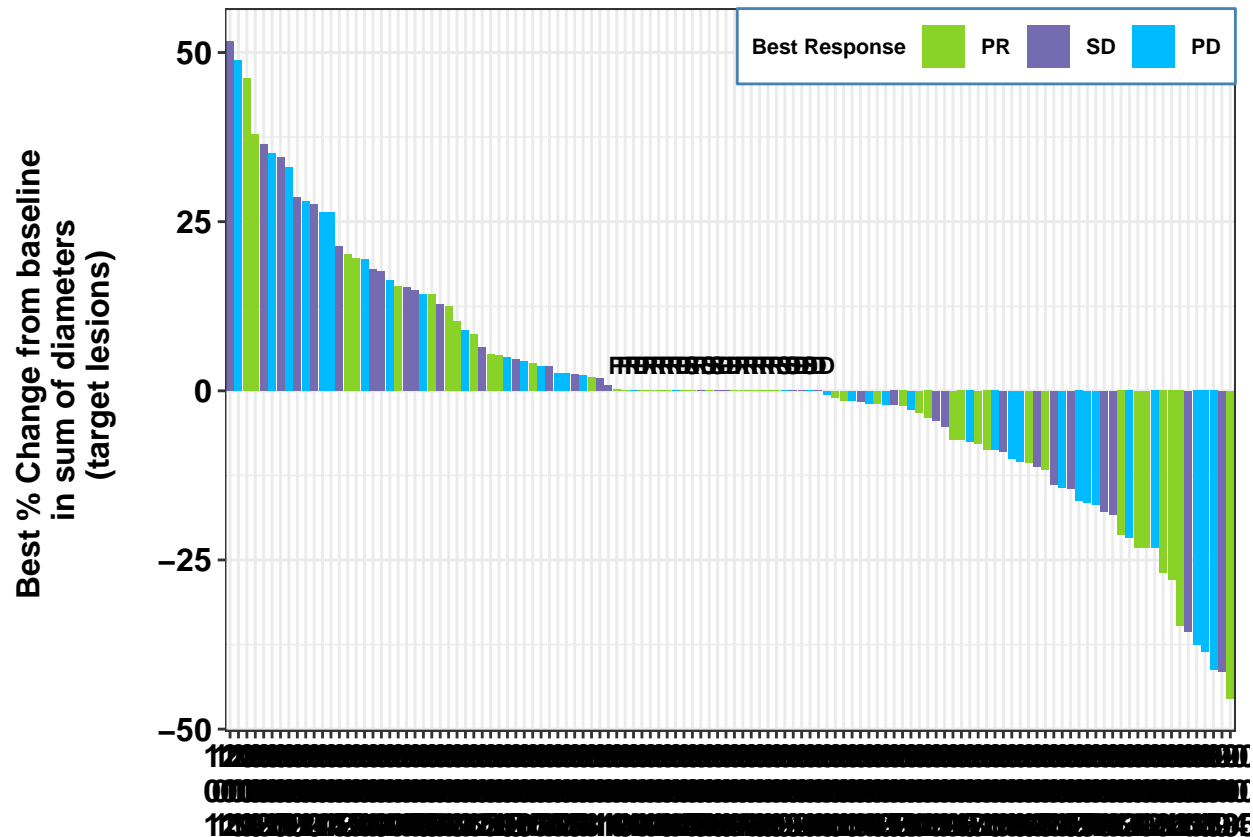
```

```

## Warning: A numeric `legend.position` argument in `theme()` was deprecated in ggplot2
## 3.5.0.
## i Please use the `legend.position.inside` argument of `theme()` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.

```

```
waterfall.plot
```



```
var <- c("DOR", "AVISIT")
var_label <- c("Duration (days)", "Timepoint")

add.plot1 <- dat_all %>%
  reshape2::melt(measure.vars = eval(var), value.name = "label", variable.name = "layer") %>%
  mutate(layer = factor(layer, levels = var, labels = var_label)) %>%
  ggplot(aes(reorder(SUBJID, -AVAL))) +
  geom_text(aes(y = layer, label = label), size = 3, fontface = "bold") +
  labs(y = "", x = NULL) +
  theme_minimal() +
  theme(
    axis.text.y = element_text(
      size = 10,
      colour = "black",
      face = "bold"
    ),
    axis.line = element_blank(),
    axis.ticks = element_blank(),
    axis.text.x = element_blank(),
    panel.grid = element_blank(),
    strip.text = element_blank()
  ) +
  coord_fixed(ratio = .8)
add.plot1
```


Duration (days)

```

footnote1 <- 'Footnote1: "\\n\" can be used to split long footnote into next line.'
footnote2 <- "Footnote2: To avoid manually adding split character. The function stringr::str_wrap  can
footnote <- lapply(c(footnote1,
                     footnote2,
                     "",
                     dat_all$IDFOOT[1]),
                  function(x) stringr::str_wrap(x, width=120)) %>% # apply str_wrap to individual foot
unlist() %>% # convert list structure to vector
stringr::str_flatten('\n') # add split character(new line) between footnotes

var <- c("ATMLOSS_L", "ATMLOSS_P", "ATMLOSS_R")
var_label <- c("Local", "Prospective", "Retrospective")

add.plot2 <- dat_all %>%
  reshape2::melt(measure.vars = eval(var),
                value.name = "label",
                variable.name = "layer") %>%
  mutate(label=case_when(
    label == "L" ~ "ATM Loss",
    label == "E" ~ "ATM Expressed",
    label == "9" ~ "Data unavailable",
    TRUE ~ label
  )) %>%
  mutate(layer = factor(layer, levels = rev(var), labels = rev(var_label))) %>%
  ggplot() +
  aes(reorder(SUBJID, -AVAL), layer, color=label, shape=label) +

```

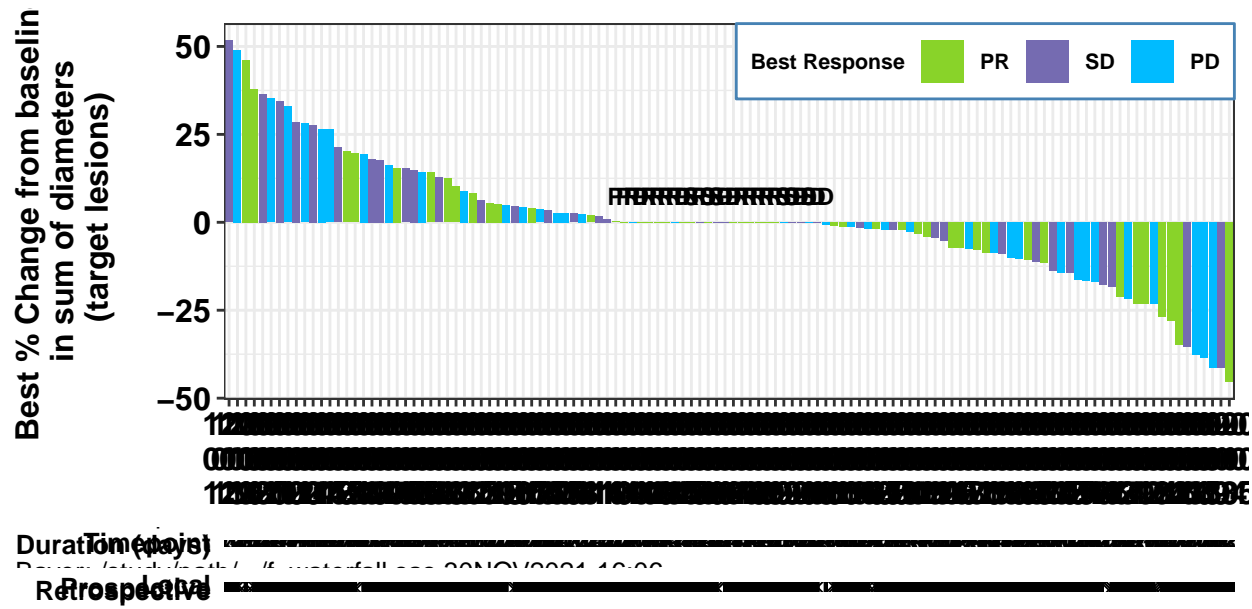
```

geom_point(size=3)+
scale_shape_manual(breaks = c("ATM Loss","ATM Expressed", "Data unavailable"),
  values = c("ATM Loss"=15,"ATM Expressed"=0,
    "Data unavailable"=7))+
scale_color_manual(values = c("ATM Loss"="black", "ATM Expressed"="black",
  "Data unavailable"= 'black'))+

theme_classic()+
theme(axis.text=element_text(size=10, colour = "black",face = "bold"),
  axis.title=element_blank(),
  axis.line = element_blank(),
  axis.ticks = element_blank(),
  axis.text.x = element_blank(),
  legend.title = element_blank(),
  legend.text = element_text(size=10, color = "black", face = "bold"),
  legend.position = "bottom",
  panel.border = element_blank(),
  panel.grid = element_blank(),
  strip.text = element_blank(),
  plot.caption = element_text(hjust = 0, size = 10),
  plot.caption.position = "plot"
)+
coord_fixed(ratio=.9)+
labs(caption = footnote)
add.plot2

p1 <- waterfall.plot %>% ggplotGrob()
p2 <- add.plot1 %>% ggplotGrob()
p3 <- add.plot2 %>% ggplotGrob()
gtable_rbind(p1, p2, p3,
  size='first') %>% grid.draw()

```



ATM Expressed ● ATM Loss ● Data unavailable ■ ATM Loss □ ATM Expressed

Footnote1: "\n" can be used to split long footnote into next line.

Footnote2: To avoid manually adding split character. The function `stringr::str_wrap` can be used here to wrap text with a specified width.

Bayer: /study/path/.../f_waterfall.sas 30NOV2021 16:06

The so called spider plot, really the clinical endpoint over time

```
# Packages
library(ggrepel)

# Data
sp <- data.frame(subjidn = rep(1:5, times = c(5,4,5,3,4)),
                 trtp = rep(c('drug', 'placebo'), times = c(8,13)),
                 avisit = c(0:4, 0:3, 0:4, 0:2, 0:3),
                 aval = c(0, 5, 15, 20, 20,
                          0, -10, -20, -15,
                          0, -30, -25, -35, -30,
                          0, 15, 30,
                          0, -20, -15, -20))

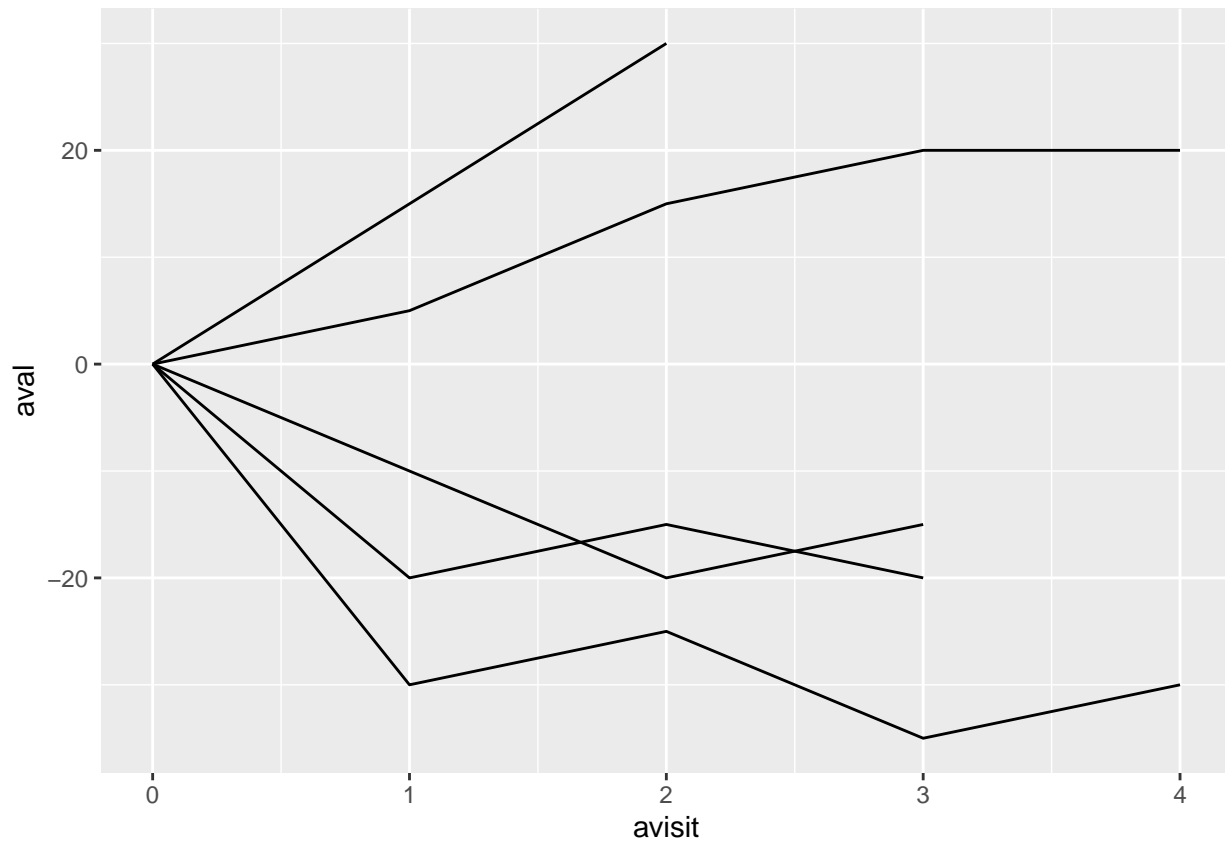
sp
```

```
##      subjidn      trtp avisit aval
## 1         1      drug        0     0
## 2         1      drug        1     5
## 3         1      drug        2    15
## 4         1      drug        3    20
## 5         1      drug        4    20
## 6         2      drug        0     0
## 7         2      drug        1    -10
## 8         2      drug        2    -20
```

```
## 9      2 placebo      3 -15
## 10     3 placebo      0  0
## 11     3 placebo      1 -30
## 12     3 placebo      2 -25
## 13     3 placebo      3 -35
## 14     3 placebo      4 -30
## 15     4 placebo      0  0
## 16     4 placebo      1 15
## 17     4 placebo      2 30
## 18     5 placebo      0  0
## 19     5 placebo      1 -20
## 20     5 placebo      2 -15
## 21     5 placebo      3 -20
```

```
basic_spider <- ggplot(sp, aes(x = avisit, y = aval, group = subjdn)) +
  geom_line()

basic_spider
```



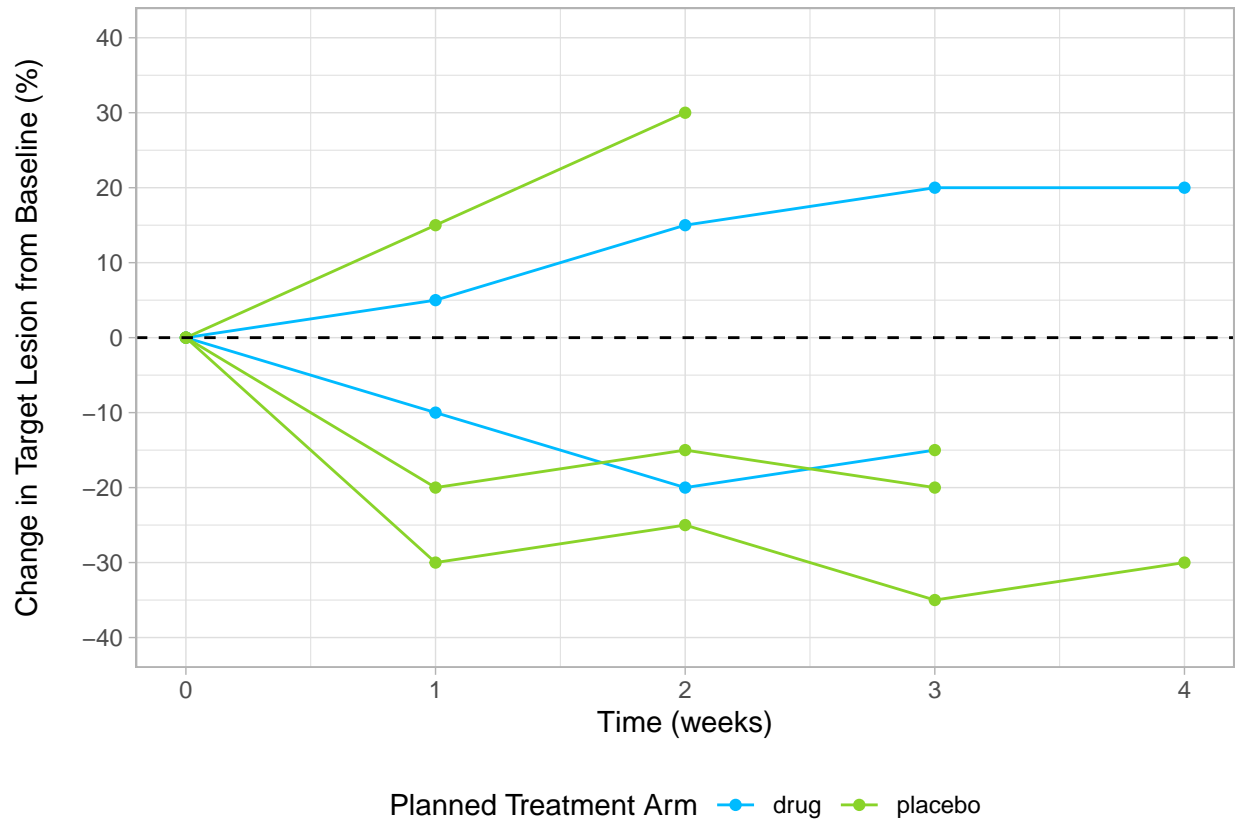
```
custom_spider <- ggplot(sp, aes(x = avisit, y = aval, group = subjdn, color = trtp)) +
  geom_line() +
  geom_point() +
  geom_hline(yintercept = 0, linetype = "dashed") +
  scale_color_manual("Planned Treatment Arm", values = c('#00bbff', '#89d329')) +
  scale_y_continuous(limits = c(-40, 40), breaks = seq(-40, 40, by = 10)) +
```

```

ylab("Change in Target Lesion from Baseline (%)") +
xlab("Time (weeks)") +
theme_light() +
theme(legend.position = "bottom")

```

custom_spider

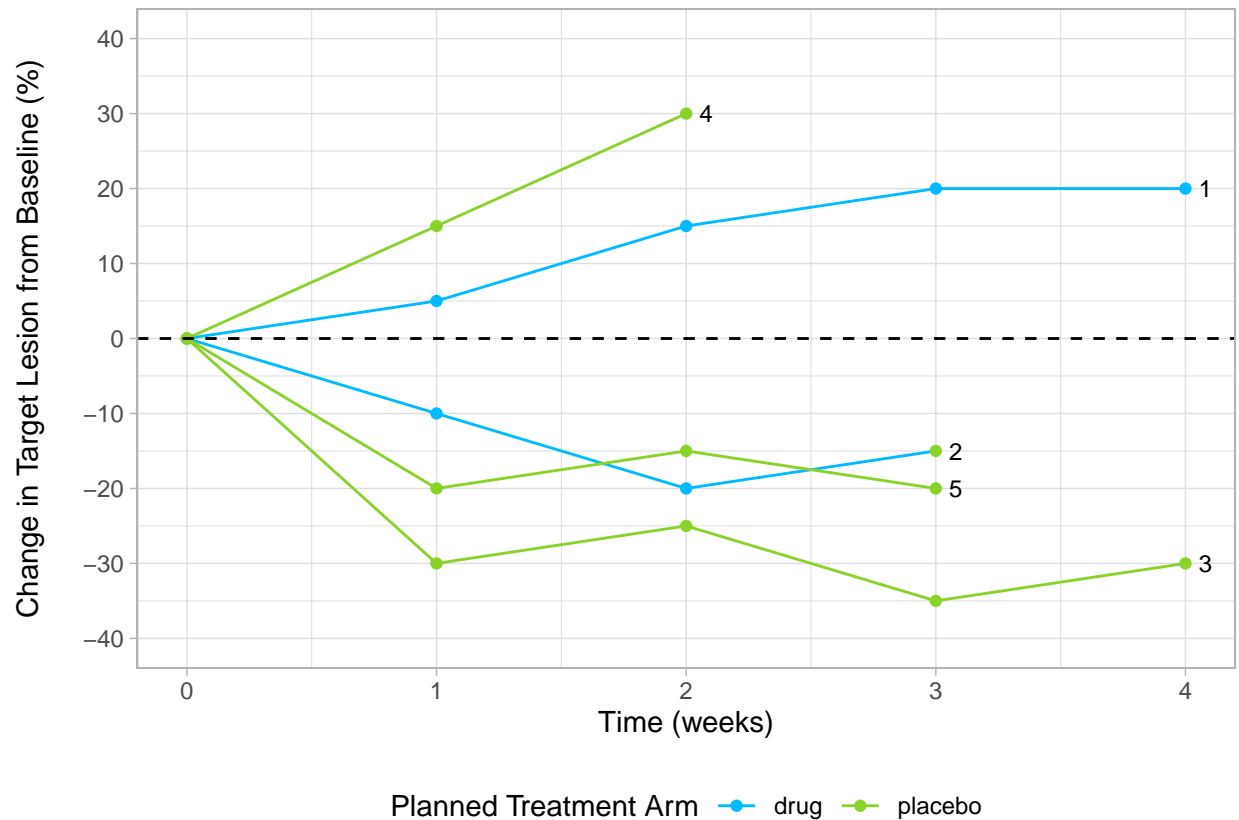


```

sp_ends <- sp %>%
  group_by(subjdn) %>%
  top_n(1, avisit)

custom_spider +
  geom_text_repel(
    aes(label = subjdn),
    color="black",
    data=sp_ends,
    size = 3,
    direction = "x",
    hjust = 1
  )

```



Now, let's look at the survival analysis and its plot

```
# Packages
library(survminer)
```

```
## Loading required package: ggpubr
```

```
##
## Attaching package: 'ggpubr'
```

```
## The following objects are masked from 'package:flextable':
##
##   border, font, rotate
```

```
library(survival)
```

```
##
## Attaching package: 'survival'
```

```
## The following object is masked from 'package:survminer':
##
##   myeloma
```

```

library(broom)
library(flextable)

# Data
adtte <- haven::read_xpt(
  paste0("https://github.com/phuse-org/TestDataFactory/",
    "raw/main/Updated/TDF_ADaM/adtte.xpt"))

surv_model <- adtte %>%
  filter(PARAMCD == "TTDE") %>%
  select(STUDYID, USUBJID, PARAMCD, AVAL, CNSR, TRTA) %>%
  survfit(Surv(AVAL, 1-CNSR) ~ TRTA, data = .)

head(tidy(surv_model))

## # A tibble: 6 x 9
##   time n.risk n.event n.censor estimate std.error conf.high conf.low strata
##   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <chr>
## 1     1     86      1      0  0.988  0.0117      1    0.966 TRTA=Plac~
## 2     2     85      1      0  0.977  0.0166      1    0.945 TRTA=Plac~
## 3     3     84      2      0  0.953  0.0238    0.999  0.910 TRTA=Plac~
## 4     7     82      1      0  0.942  0.0268    0.993  0.894 TRTA=Plac~
## 5     8     81      0      1  0.942  0.0268    0.993  0.894 TRTA=Plac~
## 6     9     80      1      0  0.930  0.0296    0.986  0.878 TRTA=Plac~

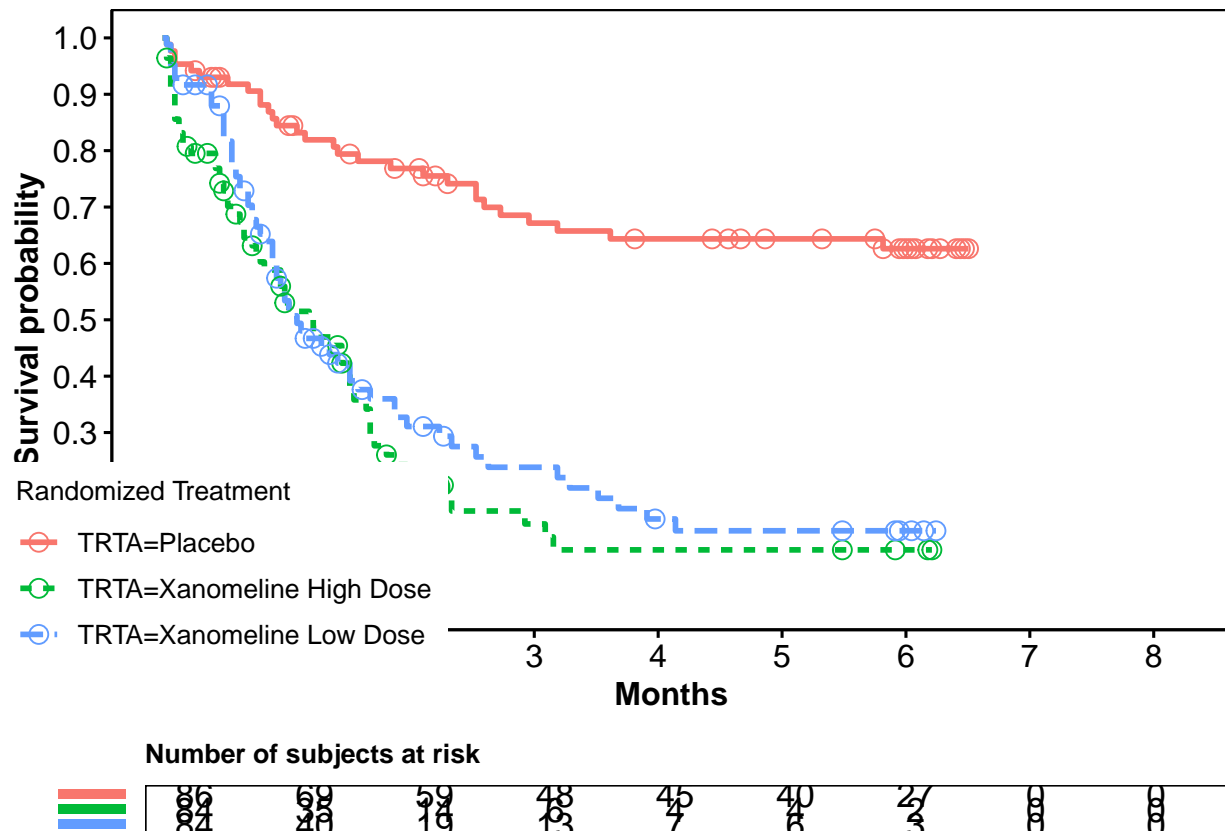
ggsurvplot(
  fit = surv_model,
  data = adtte,
  risk.table = TRUE,
  #ylab = ylabs,
  xlab = "Months",
  linetype = "strata",
  conf.int = F,
  legend.title = "Randomized Treatment",
  legend = c(0.1, 0.1),
  #palette = c(color_trt1,color_trt2),
  risk.table.title = "Number of subjects at risk",
  risk.table.y.text = F,
  risk.table.height = .15,
  censor.shape = 1,
  censor.size = 3,
  ncensor.plot = F,
  xlim = c(0,250),
  xscale = "d_m",
  break.x.by = 30.4375,
  break.y.by = .1,
  ggtheme = theme_survminer(
    font.main = c(10, "bold"),
    font.submain = c(10, "bold"),
    font.x = c(12, "bold"),
    font.y = c(12, "bold"),
  ) + theme(panel.border = element_rect(fill = NA)),

```

```

tables.theme = theme_cleantable()
)

```



```

ggsurvplot(
  fit = surv_model,
  data = adtte,
  risk.table = TRUE,
  #ylab = ylabs,
  xlab = "Months",
  linetype = "strata",
  conf.int = T,
  legend.title = "Randomized Treatment",
  legend = c(0.1, 0.1),
  #palette = c(color_trt1,color_trt2),
  risk.table.title = "Number of subjects at risk",
  risk.table.y.text = F,
  risk.table.height = .15,
  censor.shape = 1,
  censor.size = 3,
  ncensor.plot = F,
  #xlim = c(0,250),
  xscale = "d_m",
  break.x.by = 30.4375,
  break.y.by = .1,
  ggtheme = theme_survminer(

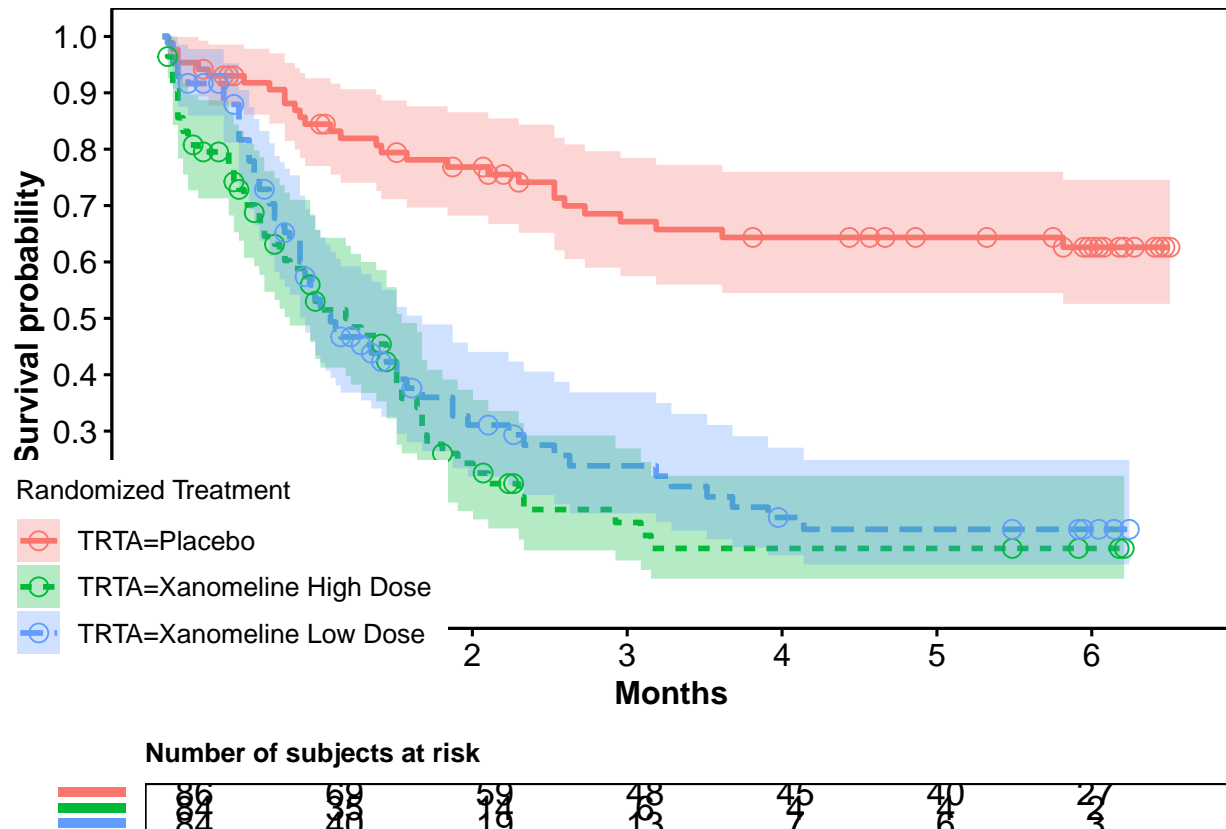
```



```

font.main = c(10, "bold"),
font.submain = c(10, "bold"),
font.x = c(12, "bold"),
font.y = c(12, "bold"),
) + theme(panel.border = element_rect(fill = NA)),
tables.theme = theme_cleantable()
)

```



FDA likes the scatter plot for safety

```

# Packages
library(ggplot2)
library(dplyr)

# Data
advs <- haven::read_xpt(
  paste0("https://github.com/phuse-org/TestDataFactory/",
    "raw/main/Updated/TDF_ADaM/advs.xpt"))

## extract information from this dataset

# This step is to obtain baseline values for each USUBJID for the parameter and analysis time point of
baseline <- advs %>%
  filter(
    SAFFL == "Y",
    PARAMCD == "SYSBP",

```

```

    ATPT == "AFTER LYING DOWN FOR 5 MINUTES",
    TRTA != "Placebo"
  ) %>%
  dplyr::select(USUBJID, TRTA, BASE) %>%
  distinct(USUBJID, .keep_all = TRUE)

# The next step is to obtain the maximum value for each USUBJID post-baseline for the same parameter and
post <- advs %>%
  filter(
    SAFFL == "Y",
    PARAMCD == "SYSBP",
    ATPT == "AFTER LYING DOWN FOR 5 MINUTES",
    TRTA != "Placebo",
    AVISIT != "Baseline",
    ANL01FL == "Y"
  ) %>%
  dplyr::select(USUBJID, AVAL) %>%
  group_by(USUBJID) %>%
  arrange(desc(AVAL)) %>%
  slice(1) %>%
  ungroup()

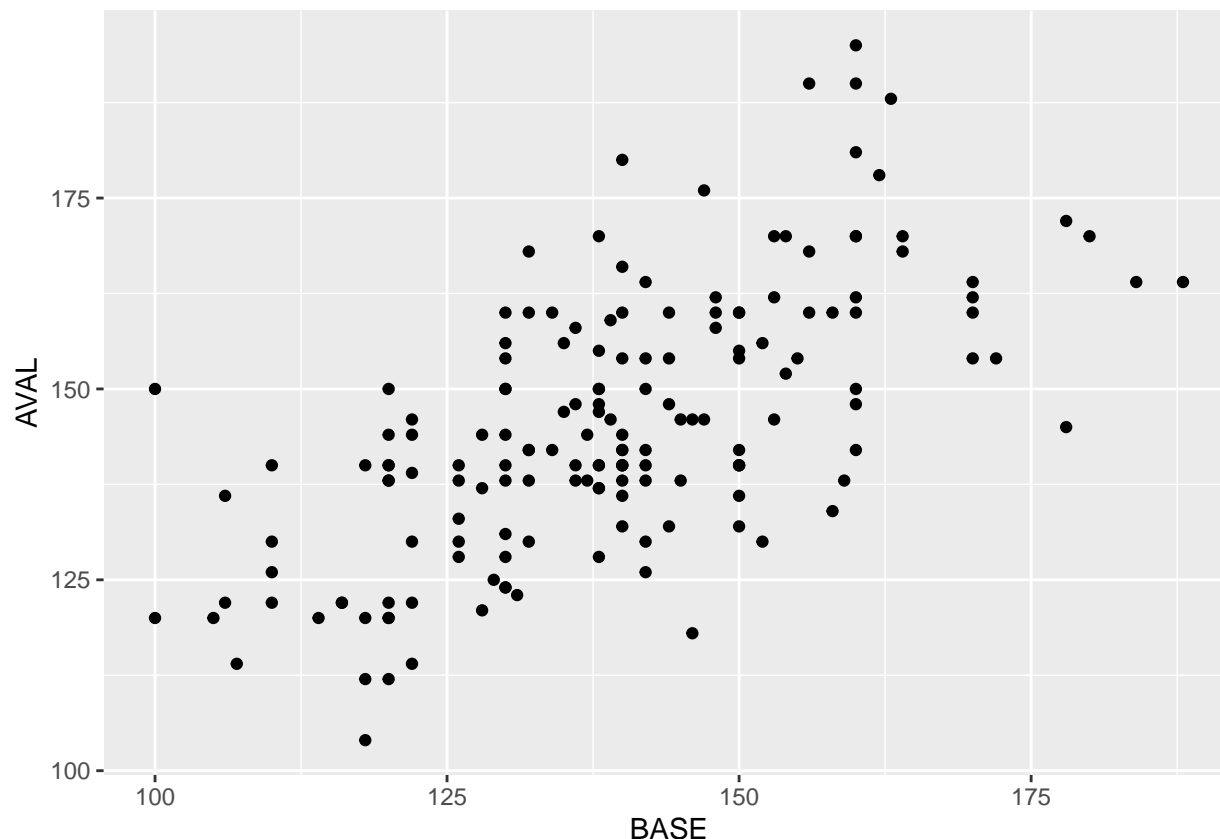
## use left_join to create the data
all <- baseline %>%
  left_join(post)

## Joining with `by = join_by(USUBJID)`

ggplot(data = all, aes(x = BASE, y = AVAL)) +
  geom_point()

## Warning: Removed 2 rows containing missing values or values outside the scale range
## (`geom_point()`).

```



```

yaxis_text <- "\nMaximum Post-baseline Systolic Blood Pressure (mmHg)\n"
xaxis_text <- "\nBaseline Systolic Blood Pressure (mmHg)"
header <- "Figure 17. Baseline vs. Maximum Systolic Blood Pressure by Treatment Arm, Safety Population"
footer1 <- "Source: [include Applicant source, datasets and/or software tools used]."
footer2 <- "'Gray dotted line = no increase; blue line = high-dose treatment linear regression; orange line = low-dose treatment linear regression'"

f17 <- ggplot(all, aes(x = BASE, y = AVAL, linetype = TRTA)) +
  geom_point(colour = "black", shape = 21, size = 4, alpha = 0.6, aes(fill = TRTA)) +
  stat_smooth(method = "lm", se = FALSE, aes(color = TRTA)) +
  scale_color_manual(values = c("skyblue", "orange")) +
  scale_fill_manual(values = c("skyblue", "orange")) +
  geom_abline(intercept = 0, slope = 1, size = 1, lty = "dotted", color = "#5A5A5A") +
  scale_x_continuous(breaks = seq(90, 210, 30), limits = c(90, 210)) +
  scale_y_continuous(breaks = seq(90, 210, 30), limits = c(90, 210)) +
  labs(
    y = yaxis_text,
    x = xaxis_text,
    title = header,
    caption = paste(footer1, footer2, sep = "\n")
  ) +
  theme_light() +
  theme(
    panel.background = element_rect(fill = NA, color = "skyblue3", size = 2, linetype = "solid"),
    plot.caption = element_text(hjust = 0),
    legend.position = "bottom",
    legend.title = element_blank(),
  )

```

```
axis.text = element_text(face = "bold")
)

## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.

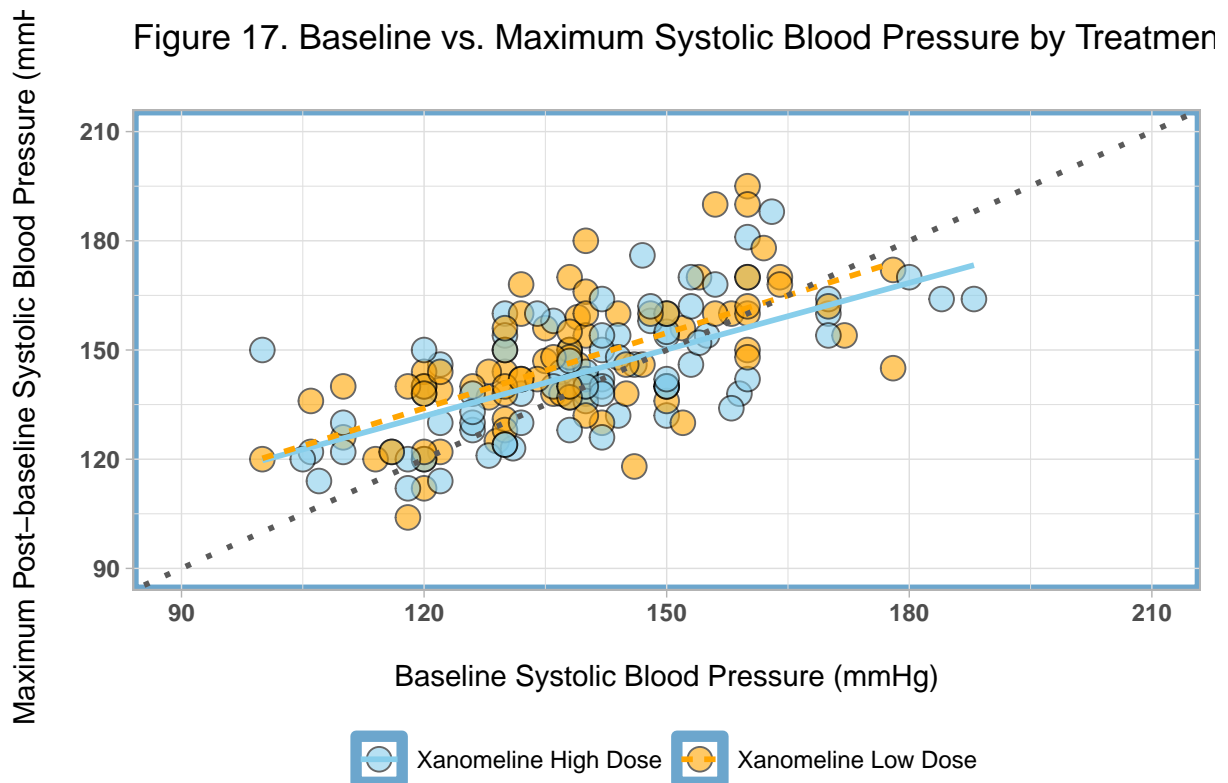
## Warning: The `size` argument of `element_rect()` is deprecated as of ggplot2 3.4.0.
## i Please use the `linewidth` argument instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.

f17

## `geom_smooth()` using formula = 'y ~ x'

## Warning: Removed 2 rows containing non-finite outside the scale range
## (`stat_smooth()`).

## Warning: Removed 2 rows containing missing values or values outside the scale range
## (`geom_point()`).
```



Source: [include Applicant source, datasets and/or software tools used].

¹Gray dotted line = no increase; blue line = high-dose treatment linear regression; orange dashed line

```
coef <- all %>%
  filter(TRTA == "Xanomeline High Dose") %>%
  lm(data = ., AVAL ~ BASE) %>%
  coefficients()
```

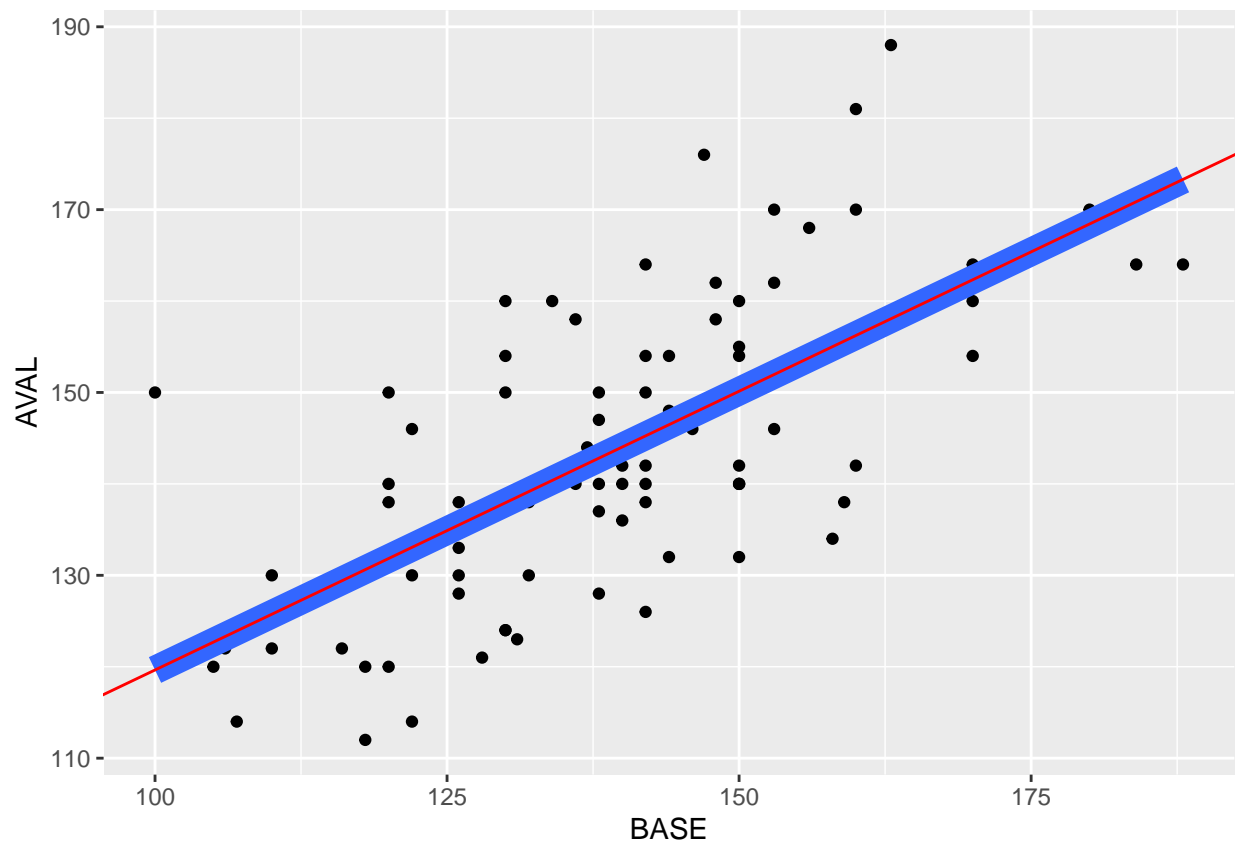
```
coef
```

```
## (Intercept)      BASE
## 58.6691789    0.6097608
```

```
all %>%
  filter(TRTA == "Xanomeline High Dose") %>%
  ggplot(., aes(x = BASE, y = AVAL)) +
  geom_point() +
  stat_smooth(method = "lm", se = FALSE, size = 5) +
  geom_abline(intercept = coef[[1]], slope = coef[[2]], color = "red")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
## Warning: Removed 2 rows containing non-finite outside the scale range (`stat_smooth()`).
## Removed 2 rows containing missing values or values outside the scale range
## (`geom_point()`).
```



```

# alternatively, plot with patchwork::plot_annotation

# setup plot text
yaxis_text_alt <- "Maximum Post-baseline Systolic Blood Pressure (mmHg)"
xaxis_text_alt <- "Baseline Systolic Blood Pressure (mmHg)"
header_alt <- "Figure 17. Baseline vs. Maximum Systolic Blood Pressure by Treatment Arm, Safety Populat
footer1_alt <- "Source: [include Applicant source, datasets and/or software tools used]."
footer2_alt <- "Gray dotted line = no increase; blue line = high-dose treatment linear regression; grey

f17_alt <- ggplot(all, aes(x = BASE, y = AVAL, linetype = TRTA)) +
  geom_point(colour = "black", shape = 21, size = 4, alpha = 0.6, aes(fill = TRTA)) +
  stat_smooth(method = "lm", se = FALSE, aes(color = TRTA)) +
  scale_color_manual(values = c("skyblue", "grey")) +
  scale_fill_manual(values = c("skyblue", "white")) +
  geom_abline(intercept = 0, slope = 1, size = 1, lty = "dotted", color = "#5A5A5A") +
  scale_x_continuous(breaks = seq(90, 210, 30), limits = c(90, 210)) +
  scale_y_continuous(breaks = seq(90, 210, 30), limits = c(90, 210)) +
  labs(
    y = yaxis_text_alt,
    x = xaxis_text_alt
  ) +
  theme_bw() +
  theme(
    plot.background = element_rect(fill = NA, color = "skyblue3", size = 2, linetype = "solid"),
    legend.position = "bottom",
    legend.title = element_blank(),
    axis.text = element_text(face = "bold")
  )

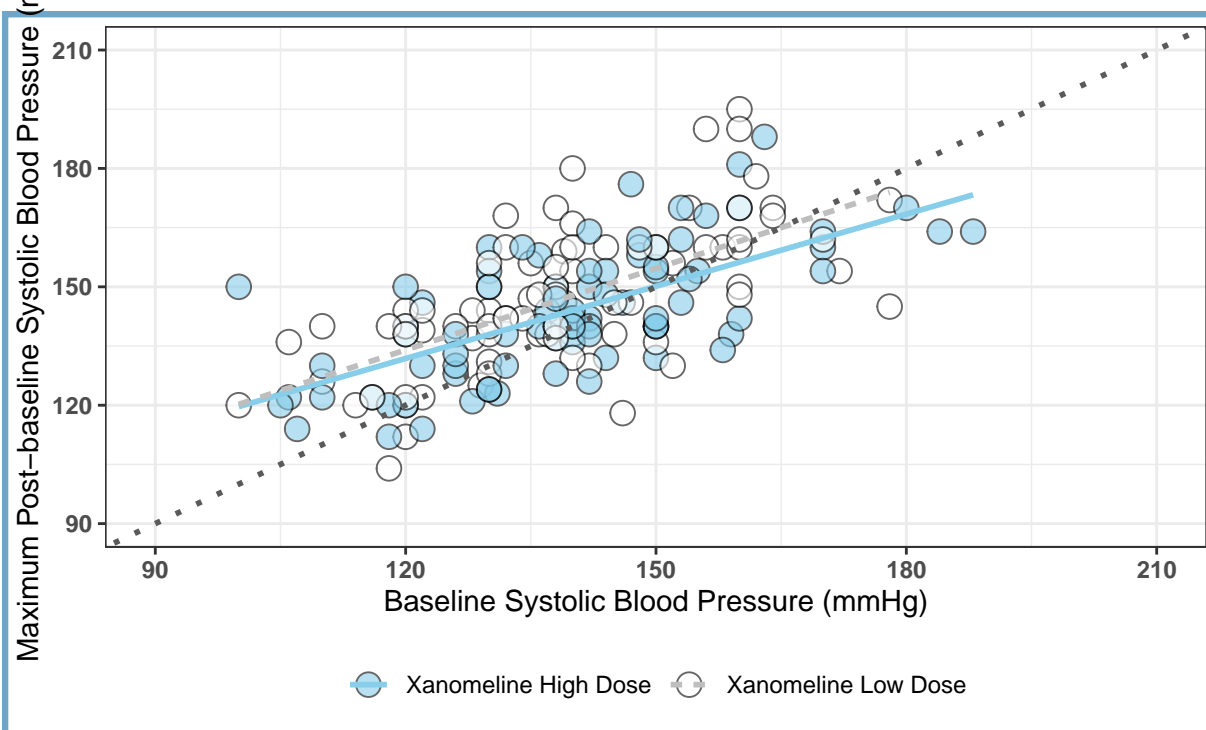
# Use the patchwork package to specify title, caption (e.g. footers) and footers position (i.e. left al
library(patchwork)
f17_alt + patchwork::plot_annotation(title = header_alt,
                                     caption = paste(footer1_alt, footer2_alt, sep = "\n"),
                                     theme = theme(plot.caption = element_text(hjust = 0)))

## `geom_smooth()` using formula = 'y ~ x'

## Warning: Removed 2 rows containing non-finite outside the scale range (`stat_smooth()`).
## Removed 2 rows containing missing values or values outside the scale range
## (`geom_point()`).

```

Figure 17. Baseline vs. Maximum Systolic Blood Pressure by Treatment Arm, S



Source: [include Applicant source, datasets and/or software tools used].

¹Gray dotted line = no increase; blue line = high-dose treatment linear regression; grey dashed line = low-dose line

ADAE data frp, PhUSE Test Data Factory repository

```
# Packages
library(ggplot2)
library(dplyr)
library(ggh4x)

# Data
adae <- haven::read_xpt(
  paste0("https://github.com/phuse-org/TestDataFactory/",
    "raw/main/Updated/TDF_ADaM/adae.xpt"))

# prepare data
start <- adae %>%
  select(TRTA, USUBJID, AESEV, AEBODSYS) %>%
  filter(TRTA != "Placebo")

event_counts <- start %>%
  distinct(USUBJID, AEBODSYS, AESEV, .keep_all = TRUE) %>%
  count(TRTA, AEBODSYS, AESEV)

ggplot(event_counts, aes(fill = AESEV, y = n, x = reorder(AEBODSYS, n))) +
  geom_bar(position="stack", stat="identity") +
  coord_flip() +
  facet_wrap(~ TRTA, scales = "free_x") +
  #facet_wrap(~ TRTA, scales = "free_y") +
  faceted_pos_scales(y = list(
```

```

scale_y_reverse(),
scale_y_continuous()),

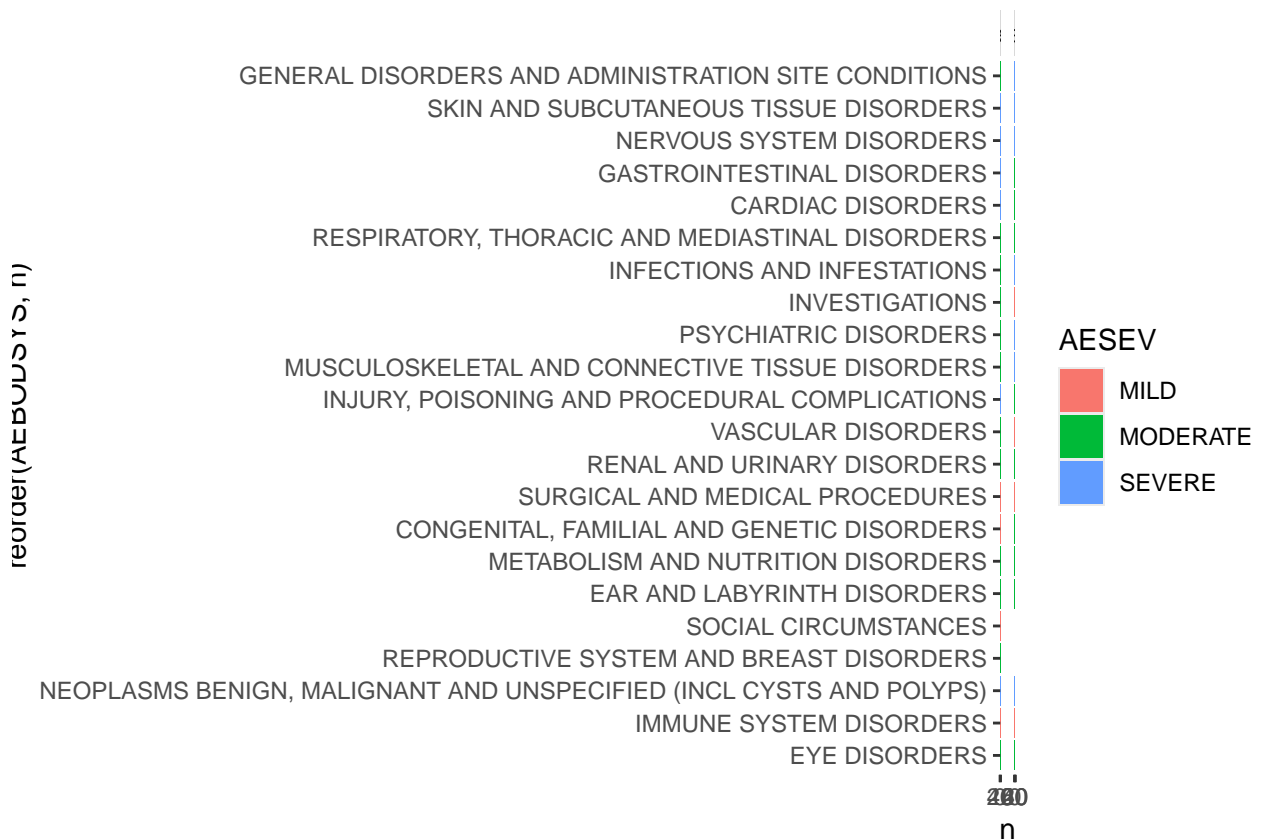
)

```

```

## Warning: Attempting to add faceted y scales, while y scales are not free.
## i Try adding `scales = "free_y"` to the facet.

```



```

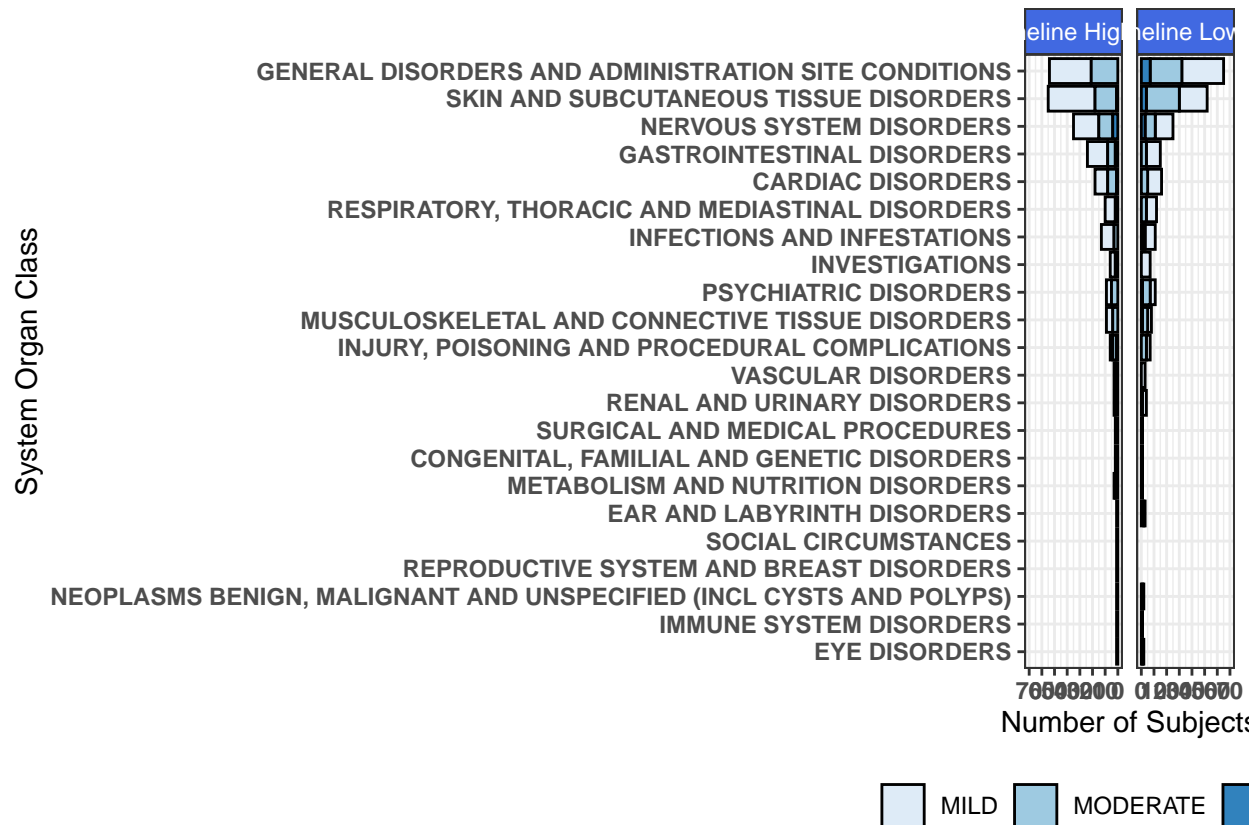
## do some customiztion

ggplot(event_counts, aes(fill = AESEV, y = n, x = reorder(AEBODSYS, n))) +
  geom_bar(position="stack", stat="identity", color = "black") +
  coord_flip() +
  scale_fill_brewer(palette = "Blues") +
  facet_wrap(~ TRTA, scales = "free_x") +
  faceted_pos_scales(y = list(
    scale_y_reverse(breaks = seq(70,0,-10), limits = c(70,0)),
    scale_y_continuous(breaks = seq(0,70,10), limits = c(0,70)))) +
  theme_bw() +
  theme(legend.position = "bottom",
        legend.title = element_blank(),
        strip.text = element_text(color = "white"),
        strip.background = element_rect(fill = "royalblue"),
        axis.text = element_text(face="bold")) +
  labs(y = "Number of Subjects", x = "System Organ Class")

```



```
## Warning: Attempting to add faceted y scales, while y scales are not free.
## i Try adding `scales = "free_y"` to the facet.
```



```
### load relevant libraries and data
```

```
library(dplyr)
library(haven)
library(patchwork)
library(ggplot2)
```

```
adam_path <- "https://github.com/phuse-org/TestDataFactory/raw/main/Updated/TDF_ADaM/"
adsl <- read_xpt(paste0(adam_path, "adsl.xpt"))
adlb <- read_xpt(paste0(adam_path, "adlbc.xpt"))
```

```
# prepare lab data set for plotting
```

```
data <- adlb %>%
  filter(PARAMCD == "SODIUM" & !is.na(AVISIT)) %>%
  select(USUBJID, TRTA, AVISIT, PARAM, AVAL, BASE, CHG) %>%
  mutate(AVISIT = factor(AVISIT, levels = c("Baseline", "Week 2", "Week 4", "Week 6", "Week 8", "Week 10",
    "Week 12", "Week 14", "Week 16", "Week 18", "Week 20", "Week 22", "Week 24", "Week 26", "End of Treatment")))
```

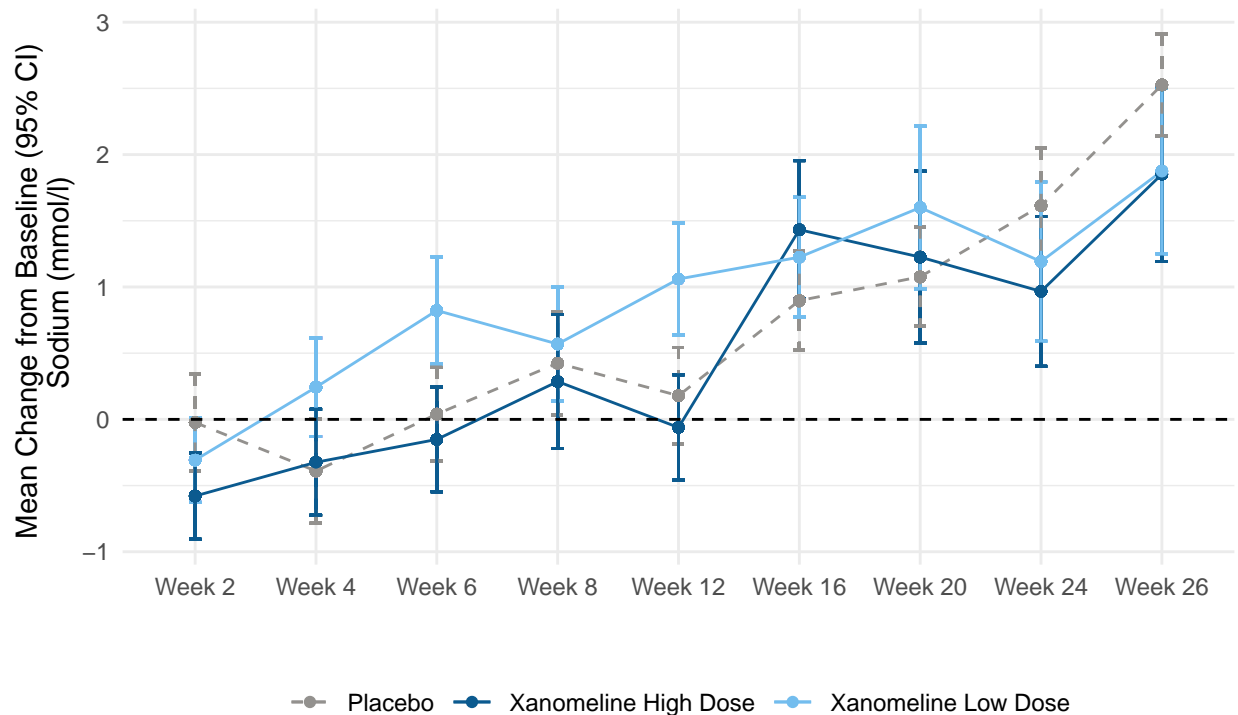
```
# create plot
```

```

p <- data %>%
  filter(!AVISIT %in% c("Baseline", "End of Treatment") & !is.na(AVISIT)) %>%
  group_by(AVISIT, TRTA) %>%
  mutate(mean_chg = mean(CHG, na.rm = TRUE),
         se_chg = sd(CHG, na.rm = TRUE) / sqrt(length(CHG))) %>%
  ungroup() %>%
  ggplot(., aes(x=AVISIT, y=mean_chg, group=TRTA, linetype = TRTA)) +
  geom_errorbar(aes(ymin=mean_chg-se_chg, ymax=mean_chg+se_chg, color=TRTA), width=.1) +
  geom_line(aes(color=TRTA)) +
  geom_point(aes(color=TRTA)) +
  geom_hline(yintercept=0, linetype='dashed')+
  scale_linetype_manual(values = c("dashed",rep("solid",2))) +
  scale_color_manual(values=c("#93918E", "#0B5A8F", "#73BDDE")) +
  labs(x = "", y = "Mean Change from Baseline (95% CI) \n Sodium (mmol/l)") +
  theme_minimal() +
  theme(legend.position="bottom",
        legend.title=element_blank()) +
  coord_fixed() +
  theme(aspect.ratio = 0.5)

```

p



```
# prepare data for table
```

```
table_data <- data %>%
```

```

filter(!AVISIT %in% c("Baseline", "End of Treatment") & !is.na(AVISIT)) %>%
group_by(AVISIT, TRTA) %>%
mutate(mean_chg = mean(CHG, na.rm = TRUE),
       mean_aval = mean(AVAL, na.rm = TRUE),
       n_patients = n_distinct(USUBJID),
       n_patients = n_distinct(USUBJID),
       TRTA = case_when(
         TRTA == "Xanomeline High Dose" ~ "High Dose",
         TRTA == "Xanomeline Low Dose" ~ "Low Dose",
         TRUE ~ TRTA)) %>%
ungroup() %>%
select(TRTA, AVISIT, mean_chg, mean_aval, n_patients) %>%
distinct()

data_table1 <- table_data %>%
  select(-n_patients) %>%
  mutate(value = paste0(round(mean_chg, 2), "/", round(mean_aval, 2)))

data_table2 <- table_data %>%
  select(-mean_chg, -mean_aval)

# plot table 1

p_table1 <- ggplot(data = data_table1, aes(x = AVISIT, y = TRTA)) +
  scale_shape_manual(values = 1:length(data_table1$TRTA))+
  ggpubr::geom_exec(geom_text, data = data_table1, label = data_table1$value, color = "TRTA", size=3) +
  theme_classic() +
  labs(title = "Mean Change from Baseline / Mean Value", x = "", y = "" ) +
  scale_color_manual(values=c("#93918E", "#0B5A8F", "#73BDEE")) +
  theme(axis.text.y = element_text(colour = c("#93918E", "#0B5A8F", "#73BDEE")),
        axis.text.x=element_blank(),
        axis.ticks.x=element_blank(),
        axis.ticks.y=element_blank()) +
  theme(legend.position="none",
        panel.border = element_rect(colour = "black", fill=NA, size=1),
        plot.title = element_text(size=10, face = "bold"))

## Warning: Vectorized input to `element_text()` is not officially supported.
## i Results may be unexpected or may change in future versions of ggplot2.

# just for the sake of displaying it

p_table1 +
  coord_fixed(ratio=.5)

```

Mean Change from Baseline / Mean Value

Placebo	-0.02/140.37/139.93/140.30/140.42/140.73/139.18/140.42/141.13/131.08/141.22/141.61/141.72/142.56
Low Dose	-0.31/139.62/140.10/140.82/140.60/140.57/140.45/141.06/141.23/141.16/141.39/141.52/142.88/142.08
High Dose	-0.58/139.09/139.32/139.65/140.06/140.29/140.46/140.06/140.16/141.57/141.23/141.80/141.57/141.85/142.41

```
## another way to plot
```

```
# plot table 2
```

```
p_table2 <- ggplot(data = data_table2, aes(x = AVISIT, y = TRTA)) +
  scale_shape_manual(values = 1:length(data_table2$TRTA)) +
  ggpubr::geom_exec(geom_text, data = data_table2, label = data_table2$n_patients, color = "TRTA", size=10) +
  theme_classic() +
  labs(title = "Number of Patients with Data", x = "", y = "" ) +
  scale_color_manual(values=c("#93918E", "#0B5A8F", "#73BDEE")) +
  theme(axis.text.y = element_text(colour = c("#93918E", "#0B5A8F", "#73BDEE")),
        axis.text.x=element_blank(),
        axis.ticks.x=element_blank(),
        axis.ticks.y=element_blank()) +
  theme(legend.position="none",
        panel.border = element_rect(colour = "black", fill=NA, size=1),
        plot.title = element_text(size=10, face = "bold"))
```

```
## Warning: Vectorized input to `element_text()` is not officially supported.
```

```
## i Results may be unexpected or may change in future versions of ggplot2.
```

```
# just for the sake of displaying it
```

```
p_table2 +
  coord_fixed(ratio=.5)
```

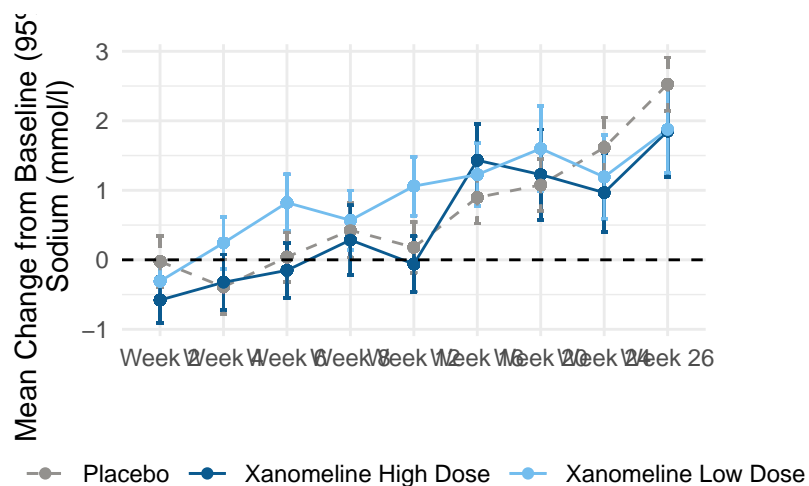
Number of Patients with Data

Placebo	84	82	75	73	67	68	65	57	57
Low Dose	80	72	64	60	52	42	31	27	25
High Dose	77	72	67	56	49	37	31	30	27

```
# combine plots

final_plot <- p + p_table1 + p_table2

# display plot w/ layout
final_plot + plot_layout(byrow = FALSE,
                        heights = c(5,1,1))
```



Mean Change from Baseline / Mean Value

Placebo	0.02/140.37	39/139.93	54/140.32	42/140.73	18/140.40	9/141.13	08/141.22	61/141.25	53/142.56
Low Dose	0.31/139.62	24/140.10	82/140.62	57/140.45	06/141.06	1.23/141.16	1.08/141.39	19/141.52	88/142.08
High Dose	0.58/139.09	32/139.65	15/140.06	29/140.46	06/140.16	43/141.57	23/141.81	97/141.57	85/142.41

Number of Patients with Data

Placebo	84	82	75	73	67	68	65	57	57
Low Dose	80	72	64	60	52	42	31	27	25
High Dose	77	72	67	56	49	37	31	30	27