Mini SDTM_DM_program

December 25, 2025

0.1 SDTM DM/SUPPDM Mini DEMO

Install and Load Package

[ ]:
```
install.packages('pacman')
```

[2]:
```
#################
pacman::p_load(haven, dplyr,tidyr, purrr, glue, lubridate,stringr, EDCimport)
```

```
Updating HTML index of packages in '.Library'

Making 'packages.html' …
 done


EDCimport installed
```

Upload the following raw datasets/exports needed to program SDTM.DM:

[3]:
```
# path <- "C:/Users/Waraba/Desktop/R in CDISC _ Jupyter Lab/rawdata"

# db <- read_all_xpt(path, format_file = NULL)  # XPT
# db <- read_all_sas(path)                      # SAS7BDAT
# db <- read_all_csv(path)                      # CSV
# load_database(db)   # puts each table into your environment

demog        <-   read_xpt("DEMOG.xpt")
ipadmin      <-   read_xpt("IPADMIN.xpt")
eos          <-   read_xpt("EOS.xpt")
enrlment     <-   read_xpt("ENRLMENT.xpt")
rand         <-   read_xpt("RAND.xpt")
box          <-   read_xpt("BOX.xpt")
adverse      <-   read_xpt("ADVERSE.xpt")
conmeds      <-   read_xpt("CONMEDS.xpt")
ecg          <-   read_xpt("ECG.xpt")
eoip         <-   read_xpt("EOIP.xpt")
eq5d3l       <-   read_xpt("EQ5D3L.xpt")
hosp         <-   read_xpt("HOSP.xpt")
```

```
lab_chem      <-   read_xpt("LAB_CHEM.xpt")
lab_hema      <-   read_xpt("LAB_HEMA.xpt")
physmeas      <-   read_xpt("PHYSMEAS.xpt")
surg          <-   read_xpt("SURG.xpt")
vitals        <-   read_xpt("VITALS.xpt")
```

Map: **Domain, studyid, subjid, siteid, usubjid, country, ethnic, race** using raw.demog

[4]:
```
demog
```

A tibble: 8 × 13

study <chr>	pt <chr>	sex <chr>	ethnic <chr>	race <chr>
STU001	1001	Male	Hispanic or Latino	White
STU001	1002	Female	Not Hispanic or Latino	Asian
STU001	1003	Male	Hispanic or Latino	Other
STU001	1004	Male	Hispanic or Latino	White
STU001	1005	Male	Not Hispanic or Latino	American Indian or Alaska Native
STU001	1006	Female	Not Hispanic or Latino	Native Hawaiian or Other Pacific Islande
STU001	1007	Male	Not Hispanic or Latino	Unknown
STU001	1008	Female	Not Hispanic or Latino	Not Reported

[5]:
```
demo1 <- demog %>%
        rename(race1=race) %>%
        mutate(
        nmiss_count = rowSums(across(c(race1, race2, race3, race4), ~ !is.na(.)␣
 ↪& . != "")),   # To count how races were selected
        race = ifelse(nmiss_count > 1, "MULTIPLE", toupper(coalesce(race1,␣
 ↪race2, race3, race4))),
        racesp = racesp,
        race1 = ifelse(nmiss_count > 1,toupper(race1),""),
        race2 = ifelse(nmiss_count > 1,toupper(race2),""),
        race3 = ifelse(nmiss_count > 1,toupper(race3),""),
        race4 = ifelse(nmiss_count > 1,toupper(race4),""),

        age = ifelse(!is.na(age_raw), as.integer(age_raw), NA_integer_),
        ageu = toupper(age_rawu),

        sex = ifelse(sex == "Female", "F", ifelse(sex == "Male", "M", sex)),
        siteid = substr(pt, 1, 2),
        usubjid = paste(study, pt, sep = "-"),

        domain = "DM",
        studyid = study,
        subjid = pt,
        country = country,
        ethnic = toupper(ethnic)
        )
```

```
demo1
```

A tibble: 8 × 22

study <chr>	pt <chr>	sex <chr>	ethnic <chr>	race1 <chr>	race2 <chr>
STU001	1001	M	HISPANIC OR LATINO		
STU001	1002	F	NOT HISPANIC OR LATINO	ASIAN	AMERICAN INDIAN C
STU001	1003	M	HISPANIC OR LATINO		
STU001	1004	M	HISPANIC OR LATINO		
STU001	1005	M	NOT HISPANIC OR LATINO		
STU001	1006	F	NOT HISPANIC OR LATINO		
STU001	1007	M	NOT HISPANIC OR LATINO		
STU001	1008	F	NOT HISPANIC OR LATINO		

Derive disposition related variables rficdtc, rfendtc, dthdtc

[6]:
```
rficdtc <- enrlment %>%
 mutate(
 rficdtc = ifelse(!is.na(icdt_raw), format(as.Date(icdt_raw, format = "%d/%b/
 ↪%Y"),"%Y-%m-%d"), NA),
 enrldtc = ifelse(!is.na(enrldt_raw), format(as.Date(enrldt_raw, format = "%d/
 ↪%b/%Y"),"%Y-%m-%d"), NA),
 randdtc = ifelse(!is.na(randdt_raw), format(as.Date(randdt_raw, format = "%d/
 ↪%b/%Y"),"%Y-%m-%d"), NA)
 ) %>%
 select(study, pt, rficdtc, enrldtc, randdtc)

rfendtc <- eos %>%
 filter(eoscat == "End of Study") %>%
 mutate(rfendtc = ifelse(!is.na(eostdt_raw), format(as.Date(eostdt_raw, format␣
 ↪= "%d/%b/%Y"),"%Y-%m-%d"), NA)) %>%
 select(study, pt, rfendtc)

dthdtc <- eos %>%
 filter(eoscat == "End of Study" & eoterm == "Death") %>%
 mutate(dthdtc = ifelse(!is.na(eostdt_raw), format(as.Date(eostdt_raw, format =␣
 ↪"%d/%b/%Y"),"%Y-%m-%d"), NA), dthfl = "Y") %>%
 select(study, pt, dthdtc, dthfl)
```

Map Exposure Related Variables: **rfxstdtc, rfxendtc**

[7]:
```
ipadmin
```

A tibble: 16 × 10

study <chr>	pt <chr>	folder <chr>	ipconc <chr>	ipstdt_raw <chr>	ipsttm_raw <chr>	ipqty_raw <chr>	ipqtyu <chr>
STU001	1004	WEEK 1	500	05/JAN/2010	8:35	2	mL
STU001	1004	WEEK 2	500	12/JAN/2010	8:35	2	mL
STU001	1004	WEEK 3	500	18/JAN/2010	9:30	1	mL
STU001	1004	WEEK 4	500	25/JAN/2010	8:45	2	mL
STU001	1005	WEEK 1	500	05/FEB/2010	8:46	2	mL
STU001	1005	WEEK 2	500	12/FEB/2010	8:30	2	mL
STU001	1005	WEEK 3	500	19/FEB/2010	8:15	0	mL
STU001	1006	WEEK 1	500	2/MAR/2010	8:30	1.9	mL
STU001	1006	WEEK 2	500	10/MAR/2010	8:30	2	mL
STU001	1007	WEEK 1	500	15/APR/2010	8:23	2	mL
STU001	1007	WEEK 2	500	22/APR/2010	9:00	2	mL
STU001	1007	WEEK 3	500	29/APR/2010	9:03	2	mL
STU001	1007	WEEK 4	500	6/MAY/2010	8:12	2	mL
STU001	1008	WEEK 1	500	27/JUN/2010	8:45	2	mL
STU001	1008	WEEK 2	500	4/JUL/2010	8:17	2	mL
STU001	1008	WEEK 3	500	11/JUL/2010	9:20	2	mL

[8]:
```
expodate1 <- ipadmin %>%
    filter(as.integer(ipqty_raw) > 0) %>%

    mutate(
    ipstdtc = as.Date(ipstdt_raw, format = "%d/%b/%Y"),
    ipsttm = format(as.POSIXct(ipsttm_raw, format = "%H:%M", tz = ""),"%H:%M"),
    infudtc = paste(ipstdtc, ipsttm, sep = "T")
 ) %>%

select(study, pt, infudtc, ipboxid)

expodate1
```

study	pt	infudtc	ipboxid
<chr>	<chr>	<chr>	<chr>
STU001	1004	2010-01-05T08:35	13434371
STU001	1004	2010-01-12T08:35	52970539
STU001	1004	2010-01-18T09:30	52120567
STU001	1004	2010-01-25T08:45	59305202
STU001	1005	2010-02-05T08:46	13787377
STU001	1005	2010-02-12T08:30	65580239
STU001	1006	2010-03-02T08:30	39024101
STU001	1006	2010-03-10T08:30	65845489
STU001	1007	2010-04-15T08:23	66223983
STU001	1007	2010-04-22T09:00	71763169
STU001	1007	2010-04-29T09:03	60038358
STU001	1007	2010-05-06T08:12	68706162
STU001	1008	2010-06-27T08:45	68891589
STU001	1008	2010-07-04T08:17	2311359
STU001	1008	2010-07-11T09:20	3199027

A tibble: 15 × 4

[9]:
```
#Earliest treatment date
rfxstdtc <- expodate1 %>%
    arrange(study,pt,infudtc) %>%
    group_by(study, pt) %>%
    slice(1) %>%
    mutate(rfxstdtc = infudtc)
rfxstdtc
```

study	pt	infudtc	ipboxid	rfxstdtc
<chr>	<chr>	<chr>	<chr>	<chr>
STU001	1004	2010-01-05T08:35	13434371	2010-01-05T08:35
STU001	1005	2010-02-05T08:46	13787377	2010-02-05T08:46
STU001	1006	2010-03-02T08:30	39024101	2010-03-02T08:30
STU001	1007	2010-04-15T08:23	66223983	2010-04-15T08:23
STU001	1008	2010-06-27T08:45	68891589	2010-06-27T08:45

A grouped_df: 5 × 5

[10]:
```
#Late treatment date
rfxendtc <- expodate1 %>%
    arrange(study,pt,infudtc) %>%
    group_by(study, pt) %>%
    slice(n()) %>%
    mutate(rfxendtc = infudtc)
rfxendtc
```

A grouped_df: 5 × 5

study	pt	infudtc	ipboxid	rfxendtc
<chr>	<chr>	<chr>	<chr>	<chr>
STU001	1004	2010-01-25T08:45	59305202	2010-01-25T08:45
STU001	1005	2010-02-12T08:30	65580239	2010-02-12T08:30
STU001	1006	2010-03-10T08:30	65845489	2010-03-10T08:30
STU001	1007	2010-05-06T08:12	68706162	2010-05-06T08:12
STU001	1008	2010-07-11T09:20	3199027	2010-07-11T09:20

Derive Planned and Actual Arm related variables

[11]:
```
enrlment
```

A tibble: 8 × 9

study	pt	folder	icdt_raw	icvers	prtvers	enrldt_raw	randdt_raw
<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>
STU001	1001	SCR	1/JAN/2010	1	1		
STU001	1002	SCR	1/JAN/2010	1	1	4/JAN/2010	
STU001	1003	SCR	1/JAN/2010	1	1	3/JAN/2010	3/JAN/2010
STU001	1004	SCR	1/JAN/2010	1	1	4/JAN/2010	5/JAN/2010
STU001	1005	SCR	15/JAN/2010	1	1	1/FEB/2010	5/FEB/2010
STU001	1006	SCR	18/FEB/2010	1	1	1/MAR/2010	1/MAR/2010
STU001	1007	SCR	4/APR/2010	2	2	14/APR/2010	14/APR/2010
STU001	1008	SCR	20/JUN/2010	2	3	26/JUN/2010	27/JUN/2010

[12]:
```
randno <- enrlment %>%
    filter(!is.na(randno) & randno!="") %>%
    select(study, pt, randno)

randno
```

A tibble: 6 × 3

study	pt	randno
<chr>	<chr>	<chr>
STU001	1003	514876
STU001	1004	101415
STU001	1005	306185
STU001	1006	987435
STU001	1007	098745
STU001	1008	123098

[13]:
```
rand
```

A tibble: 6 × 4

rand_id	tx_cd	cohort	strata
<chr>	<chr>	<chr>	<chr>
514876	PBO	1	Dummy strata1
101415	ACTIVE	1	Dummy strata1
306185	ACTIVE	1	Dummy strata1
987435	PBO	2	Dummy strata2
098745	PBO	2	Dummy strata2
123098	ACTIVE	2	Dummy strata2

[14]:
```
randtrt <- rand %>%
    mutate(
    armcd = tx_cd,
    arm = ifelse(armcd == "ACTIVE", "Active", ifelse(armcd == "PBO", "Placebo",␣
 ↪NA_character_))
) %>%
select(armcd, arm, randno=rand_id)
randtrt
```

A tibble: 6 × 3

armcd <chr>	arm <chr>	randno <chr>
PBO	Placebo	514876
ACTIVE	Active	101415
ACTIVE	Active	306185
PBO	Placebo	987435
PBO	Placebo	098745
ACTIVE	Active	123098

Merge

[15]:
```
armdata <- randno %>%
left_join(randtrt, by = "randno")

armdata
```

A tibble: 6 × 5

study <chr>	pt <chr>	randno <chr>	armcd <chr>	arm <chr>
STU001	1003	514876	PBO	Placebo
STU001	1004	101415	ACTIVE	Active
STU001	1005	306185	ACTIVE	Active
STU001	1006	987435	PBO	Placebo
STU001	1007	098745	PBO	Placebo
STU001	1008	123098	ACTIVE	Active

Derive actual ARM related variable

[16]:
```
actarmcode <- rfxstdtc
actarmcode
```

A grouped_df: 5 × 5

study <chr>	pt <chr>	infudtc <chr>	ipboxid <chr>	rfxstdtc <chr>
STU001	1004	2010-01-05T08:35	13434371	2010-01-05T08:35
STU001	1005	2010-02-05T08:46	13787377	2010-02-05T08:46
STU001	1006	2010-03-02T08:30	39024101	2010-03-02T08:30
STU001	1007	2010-04-15T08:23	66223983	2010-04-15T08:23
STU001	1008	2010-06-27T08:45	68891589	2010-06-27T08:45

Box data for mapping actual arm

[17]:
```
box
```

A tibble: 16 × 2

kitid <chr>	content <chr>
13434371	ACTIVE
52970539	ACTIVE
52120567	ACTIVE
59305202	ACTIVE
13787377	PBO
65580239	ACTIVE
45377264	ACTIVE
39024101	PBO
65845489	PBO
66223983	PBO
71763169	PBO
60038358	PBO
68706162	PBO
68891589	ACTIVE
2311359	ACTIVE
3199027	ACTIVE

[18]:
```
boxdata <- box %>%
    mutate(
    ipboxid = kitid,
    actarmcd = case_when(
        content == "ACTIVE" ~ "ACTIVE",
        content == "PBO" ~ "PBO",
        TRUE ~ NA_character_
     ),

    actarm = case_when(
        content == "ACTIVE" ~ "Active",
        content == "PBO" ~ "Placebo",
        TRUE ~ NA_character_)
)
boxdata
```

A tibble: 16 × 5

kitid <chr>	content <chr>	ipboxid <chr>	actarmcd <chr>	actarm <chr>
13434371	ACTIVE	13434371	ACTIVE	Active
52970539	ACTIVE	52970539	ACTIVE	Active
52120567	ACTIVE	52120567	ACTIVE	Active
59305202	ACTIVE	59305202	ACTIVE	Active
13787377	PBO	13787377	PBO	Placebo
65580239	ACTIVE	65580239	ACTIVE	Active
45377264	ACTIVE	45377264	ACTIVE	Active
39024101	PBO	39024101	PBO	Placebo
65845489	PBO	65845489	PBO	Placebo
66223983	PBO	66223983	PBO	Placebo
71763169	PBO	71763169	PBO	Placebo
60038358	PBO	60038358	PBO	Placebo
68706162	PBO	68706162	PBO	Placebo
68891589	ACTIVE	68891589	ACTIVE	Active
2311359	ACTIVE	2311359	ACTIVE	Active
3199027	ACTIVE	3199027	ACTIVE	Active

Merge 'actarmcode' and 'boxdata' data frames by 'ipboxid'

[19]:
```
actarmdata <- left_join(
                      actarmcode, boxdata, by = "ipboxid") %>%
              filter(!is.na(actarmcd)) %>%
              select(study, pt, actarmcd, actarm)

actarmdata
```

A grouped_df: 5 × 4

study <chr>	pt <chr>	actarmcd <chr>	actarm <chr>
STU001	1004	ACTIVE	Active
STU001	1005	PBO	Placebo
STU001	1006	PBO	Placebo
STU001	1007	PBO	Placebo
STU001	1008	ACTIVE	Active

Reference End of participation

Combine the raw date variables into 'combdate' data frame

[20]:
```
hospc <-hosp %>% mutate(
    study = as.character(study))
```

[21]:
```
combdate <- bind_rows(
 adverse %>% select(study, pt, date = aestdt_raw),
 adverse %>% select(study, pt, date = aeendt_raw),
 adverse %>% select(study, pt, date = hadmtdt_raw),
 adverse %>% select(study, pt, date = hdsdt_raw),
 conmeds %>% select(study, pt, date = cmstdt_raw),
 conmeds %>% select(study, pt, date = cmendt_raw),
```

```
 ecg %>% select(study, pt, date = egdt_raw),
 enrlment %>% select(study, pt, date = icdt_raw),
 enrlment %>% select(study, pt, date = enrldt_raw),
 enrlment %>% select(study, pt, date = randdt_raw),

 eos %>% select(study, pt, date = eostdt_raw),
 eoip %>% select(study, pt, date = eostdt_raw),
 eq5d3l %>% select(study, pt, date = dt_raw),


 ipadmin %>% select(study, pt, date = ipstdt_raw),
 lab_chem %>% select(study, pt, date = lbdt_raw),
 lab_hema %>% select(study, pt, date = lbdt_raw),
 physmeas %>% select(study, pt, date = pmdt_raw),
 surg %>% select(study, pt, date = surgdt_raw),
 vitals %>% select(study, pt, date = vsdt_raw)
)
combdate
```

study	pt	date
<chr>	<chr>	<chr>
STU001	1001	01/JAN/2010
STU001	1003	05/JAN/2010
STU001	1004	01/JAN/2010
STU001	1004	03/JAN/2010
STU001	1004	08/JAN/2010
STU001	1004	10/JAN/2010
STU001	1005	18/FEB/2010
STU001	1006	UN/MAR/2010
STU001	1007	9/MAY/2010
STU001	1001	01/JAN/2010
STU001	1003	05/JAN/2010
STU001	1004	01/JAN/2010
STU001	1004	07/JAN/2010
STU001	1004	09/JAN/2010
STU001	1004	
STU001	1005	21/FEB/2010
STU001	1006	25/MAR/2010
STU001	1007	12/MAY/2010
STU001	1001	
STU001	1003	5/JAN/2010
STU001	1004	
STU001	1004	
STU001	1004	
STU001	1004	
STU001	1005	20/FEB/2020
STU001	1006	
STU001	1007	
STU001	1001	
STU001	1003	5/JAN/2010

A tibble: 953 × 3

STU001	1004	
STU001	1008	11/Jul/2010
STU001	1008	11/Jul/2010
STU001	1008	11/Jul/2010
STU001	1008	11/Jul/2010
STU001	1008	11/Jul/2010
STU001	1008	11/Jul/2010
STU001	1008	11/Jul/2010
STU001	1008	11/Jul/2010
STU001	1008	11/Jul/2010
STU001	1008	11/Jul/2010
STU001	1008	11/Jul/2010
STU001	1008	11/Jul/2010
STU001	1008	11/Jul/2010
STU001	1008	11/Jul/2010
STU001	1008	11/Jul/2010
STU001	1008	11/Jul/2010
STU001	1004	4/JAN/2010
STU001	1005	5/FEB/2010
STU001	1006	1/MAR/2010
STU001	1007	13/APR/2010

Process the date variables to create date in ISO format in 'alldates02' data frame

[22]:
```
combdate01 <- combdate %>%
         mutate(
             dayn = suppressWarnings(as.integer(stringr::word(date, 1, sep = "/
 ↪"))),
             day = if_else(!is.na(dayn), sprintf("%02d", dayn), "-"),
             monthc = toupper(word(date, 2, sep='/')),
             month = case_when(
             monthc == "JAN" ~ "01",
             monthc == "FEB" ~ "02",
             monthc == "MAR" ~ "03",
             monthc == "APR" ~ "04",
             monthc == "MAY" ~ "05",
             monthc == "JUN" ~ "06",
             monthc == "JUL" ~ "07",
             monthc == "AUG" ~ "08",
             monthc == "SEP" ~ "09",
             monthc == "OCT" ~ "10",
             monthc == "NOV" ~ "11",
             monthc == "DEC" ~ "12",
             TRUE ~ "-"
         ),

         year = word(date,3,sep='/'),
         year = if_else(toupper(year) == "UNK", "-", year),

         datec = str_c(year, month, day, sep = "-"),
         datec = ifelse(str_sub(datec, -5) == "-----", str_sub(datec, end =␣
 ↪-6), datec),
         datec = ifelse(str_sub(datec, -4) == "----",  str_sub(datec, end =␣
 ↪-5), datec),
         datec = ifelse(str_sub(datec, -2) == "--",    str_sub(datec, end =␣
 ↪-3), datec)
)

combdate01
```

A tibble: 953 × 9

study	pt	date	dayn	day	monthc	month	year	datec
<chr>	<chr>	<chr>	<int>	<chr>	<chr>	<chr>	<chr>	<chr>
STU001	1001	01/JAN/2010	1	01	JAN	01	2010	2010-01-01
STU001	1003	05/JAN/2010	5	05	JAN	01	2010	2010-01-05
STU001	1004	01/JAN/2010	1	01	JAN	01	2010	2010-01-01
STU001	1004	03/JAN/2010	3	03	JAN	01	2010	2010-01-03
STU001	1004	08/JAN/2010	8	08	JAN	01	2010	2010-01-08
STU001	1004	10/JAN/2010	10	10	JAN	01	2010	2010-01-10
STU001	1005	18/FEB/2010	18	18	FEB	02	2010	2010-02-18
STU001	1006	UN/MAR/2010	NA	-	MAR	03	2010	2010-03
STU001	1007	9/MAY/2010	9	09	MAY	05	2010	2010-05-09
STU001	1001	01/JAN/2010	1	01	JAN	01	2010	2010-01-01
STU001	1003	05/JAN/2010	5	05	JAN	01	2010	2010-01-05
STU001	1004	01/JAN/2010	1	01	JAN	01	2010	2010-01-01
STU001	1004	07/JAN/2010	7	07	JAN	01	2010	2010-01-07
STU001	1004	09/JAN/2010	9	09	JAN	01	2010	2010-01-09
STU001	1004		NA	-	NA	-	NA	NA
STU001	1005	21/FEB/2010	21	21	FEB	02	2010	2010-02-21
STU001	1006	25/MAR/2010	25	25	MAR	03	2010	2010-03-25
STU001	1007	12/MAY/2010	12	12	MAY	05	2010	2010-05-12
STU001	1001		NA	-	NA	-	NA	NA
STU001	1003	5/JAN/2010	5	05	JAN	01	2010	2010-01-05
STU001	1004		NA	-	NA	-	NA	NA
STU001	1004		NA	-	NA	-	NA	NA
STU001	1004		NA	-	NA	-	NA	NA
STU001	1004		NA	-	NA	-	NA	NA
STU001	1005	20/FEB/2020	20	20	FEB	02	2020	2020-02-20
STU001	1006		NA	-	NA	-	NA	NA
STU001	1007		NA	-	NA	-	NA	NA
STU001	1001		NA	-	NA	-	NA	NA
STU001	1003	5/JAN/2010	5	05	JAN	01	2010	2010-01-05
STU001	1004		NA	-	NA	-	NA	NA
STU001	1008	11/Jul/2010	11	11	JUL	07	2010	2010-07-11
STU001	1008	11/Jul/2010	11	11	JUL	07	2010	2010-07-11
STU001	1008	11/Jul/2010	11	11	JUL	07	2010	2010-07-11
STU001	1008	11/Jul/2010	11	11	JUL	07	2010	2010-07-11
STU001	1008	11/Jul/2010	11	11	JUL	07	2010	2010-07-11
STU001	1008	11/Jul/2010	11	11	JUL	07	2010	2010-07-11
STU001	1008	11/Jul/2010	11	11	JUL	07	2010	2010-07-11
STU001	1008	11/Jul/2010	11	11	JUL	07	2010	2010-07-11
STU001	1008	11/Jul/2010	11	11	JUL	07	2010	2010-07-11
STU001	1008	11/Jul/2010	11	11	JUL	07	2010	2010-07-11
STU001	1008	11/Jul/2010	11	11	JUL	07	2010	2010-07-11
STU001	1008	11/Jul/2010	11	11	JUL	07	2010	2010-07-11
STU001	1008	11/Jul/2010	11	11	JUL	07	2010	2010-07-11
STU001	1008	11/Jul/2010	11	11	JUL	07	2010	2010-07-11
STU001	1008	11/Jul/2010	11	11	JUL	07	2010	2010-07-11
STU001	1008	11/Jul/2010	11	11	JUL	07	2010	2010-07-11
STU001	1004	4/JAN/2010	4	04	JAN	01	2010	2010-01-04
STU001	1005	5/FEB/2010	5	05	FEB	02	2010	2010-02-05
STU001	1006	1/MAR/2010	1	01	MAR	03	2010	2010-03-01
STU001	1007	13/APR/2010	13	13	APR	04	2010	2010-04-13

Pick the latest non-missing date for each subject

[23]:
```
rfpendtc <- combdate01 %>%
          filter(!is.na(datec) & datec != "") %>%
          arrange(study, pt, datec) %>%
          group_by(study, pt) %>%
          slice(n()) %>%
          ungroup() %>%

          select(study, pt, rfpendtc = datec)
rfpendtc
```

A tibble: 8 × 3

study <chr>	pt <chr>	rfpendtc <chr>
STU001	1001	2010-01-01
STU001	1002	2010-01-05
STU001	1003	2010-01-05
STU001	1004	2010-02-28
STU001	1005	2020-02-20
STU001	1006	2010-03-25
STU001	1007	2010-06-12
STU001	1008	2010-08-18

[ ]:
```

```

Merge all datasets together

[24]:
```
demo2 <- demo1 %>%
              left_join(rficdtc, by = c("study", "pt")) %>%
              left_join(dthdtc, by = c("study", "pt")) %>%
              left_join(rfendtc, by = c("study", "pt")) %>%
              left_join(rfxstdtc, by = c("study", "pt")) %>%
              left_join(rfxendtc, by = c("study", "pt")) %>%
              left_join(actarmdata, by = c("study", "pt")) %>%
              left_join(armdata, by = c("study", "pt")) %>%
              left_join(rfpendtc, by = c("study", "pt"))

demo2
```

A tibble: 8 × 40

study <chr>	pt <chr>	sex <chr>	ethnic <chr>	race1 <chr>	race2 <chr>
STU001	1001	M	HISPANIC OR LATINO		
STU001	1002	F	NOT HISPANIC OR LATINO	ASIAN	AMERICAN INDIAN
STU001	1003	M	HISPANIC OR LATINO		
STU001	1004	M	HISPANIC OR LATINO		
STU001	1005	M	NOT HISPANIC OR LATINO		
STU001	1006	F	NOT HISPANIC OR LATINO		
STU001	1007	M	NOT HISPANIC OR LATINO		
STU001	1008	F	NOT HISPANIC OR LATINO		

Derive additional variables that depend on previously derived variables.

[26]:
```
demo3 <- demo2 %>%
         mutate(
             rfstdtc = substr(rfxstdtc, 1, 10),
             rfstdtc = ifelse(is.na(rfstdtc) & !is.na(randdtc), randdtc,␣
 ↪rfstdtc),
             rfstdtc = ifelse(is.na(rfstdtc) & !is.na(rficdtc), rficdtc,␣
 ↪rfstdtc),

             armcd = case_when(
                         is.na(enrldtc) ~ "SCRNFAIL",
                         is.na(randdtc) ~ "NOTASSGN",
                         TRUE ~ armcd),
             arm =   case_when(
                         armcd =="SCRNFAIL" ~ "Screen Failure",
                         armcd == "NOTASSGN" ~ "Not Assigned",
                         TRUE ~ arm),
             actarmcd = case_when(
                         is.na(enrldtc) ~ "SCRNFAIL",
                         is.na(randdtc) ~ "NOTASSGN",
                         is.na(rfxstdtc) ~ "NOTTRT",
                         TRUE ~ actarmcd),

             actarm = case_when(
                     actarmcd =="SCRNFAIL" ~ "Screen Failure",
                     actarmcd == "NOTASSGN" ~ "Not Assigned",
                     actarmcd == "NOTTRT" ~ "Not Treated",
                     TRUE ~ actarm)) %>%
 rename_all(toupper)

demo3
```

A tibble: 8 × 41

STUDY <chr>	PT <chr>	SEX <chr>	ETHNIC <chr>	RACE1 <chr>	RACE2 <chr>
STU001	1001	M	HISPANIC OR LATINO		
STU001	1002	F	NOT HISPANIC OR LATINO	ASIAN	AMERICAN INDIAN
STU001	1003	M	HISPANIC OR LATINO		
STU001	1004	M	HISPANIC OR LATINO		
STU001	1005	M	NOT HISPANIC OR LATINO		
STU001	1006	F	NOT HISPANIC OR LATINO		
STU001	1007	M	NOT HISPANIC OR LATINO		
STU001	1008	F	NOT HISPANIC OR LATINO		

Write attributes and keep only required variables and in the required order

[27]:
```
varlist <- c(
```

```
 'STUDYID', 'DOMAIN', 'USUBJID', 'SUBJID', 'RFSTDTC', 'RFENDTC', 'RFXSTDTC',␣
 ↪'RFXENDTC',
 'RFICDTC', 'RFPENDTC', 'DTHDTC', 'DTHFL', 'SITEID', 'AGE', 'AGEU', 'SEX',␣
 ↪'RACE', 'ETHNIC',
 'ARMCD', 'ARM', 'ACTARMCD', 'ACTARM', 'COUNTRY', 'RACE1', 'RACE2', 'RACE3',␣
 ↪'RACE4', 'RACESP'
)
```

[29]:
```
dm <- demo3 %>%
 select(all_of(varlist))

dm
```

A tibble: 8 × 28

STUDYID <chr>	DOMAIN <chr>	USUBJID <chr>	SUBJID <chr>	RFSTDTC <chr>	RFENDTC <chr>	RFXSTDTC <chr>
STU001	DM	STU001-1001	1001	2010-01-01	NA	NA
STU001	DM	STU001-1002	1002	2010-01-01	2010-01-05	NA
STU001	DM	STU001-1003	1003	2010-01-03	2010-01-05	NA
STU001	DM	STU001-1004	1004	2010-01-05	2010-02-28	2010-01-05T08:3
STU001	DM	STU001-1005	1005	2010-02-05	NA	2010-02-05T08:4
STU001	DM	STU001-1006	1006	2010-03-02	2010-03-25	2010-03-02T08:3
STU001	DM	STU001-1007	1007	2010-04-15	2010-06-12	2010-04-15T08:2
STU001	DM	STU001-1008	1008	2010-06-27	2010-08-18	2010-06-27T08:4

[30]:
```
write_xpt(dm, "dm.xpt")
```

[31]:
```
colnames(dm)
```

1. 'STUDYID' 2. 'DOMAIN' 3. 'USUBJID' 4. 'SUBJID' 5. 'RFSTDTC' 6. 'RFENDTC' 7. 'RFXSTDTC' 8. 'RFXENDTC' 9. 'RFICDTC' 10. 'RFPENDTC' 11. 'DTHDTC' 12. 'DTHFL' 13. 'SITEID' 14. 'AGE' 15. 'AGEU' 16. 'SEX' 17. 'RACE' 18. 'ETHNIC' 19. 'ARMCD' 20. 'ARM' 21. 'ACTARMCD' 22. 'ACTARM' 23. 'COUNTRY' 24. 'RACE1' 25. 'RACE2' 26. 'RACE3' 27. 'RACE4' 28. 'RACESP'

[ ]:
```

```

[35]:
```
suppdm <- dm %>%
  select(STUDYID, USUBJID, RACE1, RACE2, RACESP) %>%
  pivot_longer(
    cols = c(RACE1, RACE2, RACESP),
    names_to = "QNAM",
    values_to = "QVAL"
  ) %>%

  filter(!is.na(QVAL) & QVAL != "") %>%
  mutate(
    RDOMAIN   = "DM",
```

```
    IDVAR     = "",
    IDVARVAL  = "",
    QNAM      = str_to_upper(str_sub(QNAM, 1, 8)),
    QLABEL    = case_when(
      QNAM == "RACE1"   ~ "Race Component 1",
      QNAM == "RACE2"   ~ "Race Component 2",
      QNAM == "RACESP" ~ "Race, Specify",
      TRUE              ~ QNAM
    ),
    QVAL   = as.character(QVAL),
    QORIG  = case_when(
      QNAM %in% c("RACE1","RACE2","RACESP") ~ "CRF",
      TRUE ~ "DERIVED"
    ),
    QEVAL = ""
  ) %>%
  select(STUDYID, RDOMAIN, USUBJID, IDVAR, IDVARVAL, QNAM, QLABEL, QVAL, QORIG,␣
 ↪QEVAL) %>%
  arrange(STUDYID, USUBJID, QNAM)

suppdm
```

A tibble: 3 × 10

STUDYID <chr>	RDOMAIN <chr>	USUBJID <chr>	IDVAR <chr>	IDVARVAL <chr>	QNAM <chr>	QLABEL <chr>
STU001	DM	STU001-1002			RACE1	Race Component
STU001	DM	STU001-1002			RACE2	Race Component
STU001	DM	STU001-1003			RACESP	Race, Specify

[ ]: