Patient-Days Table

Steph Reynolds (Stephanie.Reynolds@ucsf.edu)

Dec 15, 2021 09:19 AM

Background

Project

MINDSCAPE: Modeling of infectious network dynamics for surveillance, control and prevention enhancement

Description

This file parses the admit and discharge date from the `dm` dataset and returns an expanded dataset with one row per patient per day of hospitalization.

Source Data

- Demographics and Events Table (`dm_covid.csv`)
  - This file contains data on patient demographics (age, sex, race, ethnicity), time of admission and discharge, time of death (if applicable), length of stay (LOS), and other comorbidities.

Load required packages

```
library(here)
library(tidyverse)
  #library(forcats)
```

Import and preview data

```
## Rows: 1117 Columns: 52
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (45): deid_enc_id, deid_pat_id, sex, zip_code, pat_race, ethnicity, mar...
## dbl   (3): age, BMI, hospital_LOS
## dttm  (4): covid_pos_time, encounter_start_time, admit_time, discharge_time
```

```
##
## 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.

## spec_tbl_df [1,117 x 52] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ deid_enc_id               : chr [1:1117] "73cb518a-8700-4533-9cb4-536853a2a953" "8311e4f5-c0f4-40
##  $ deid_pat_id               : chr [1:1117] "0d2fe3af-609f-40ce-ab46-c501dc85877b" "6bbc280c-f747-49
##  $ age                       : num [1:1117] 38 63 37 71 51 33 19 33 50 30 ...
##  $ sex                       : chr [1:1117] "Male" "Female" "Male" "Female" ...
##  $ zip_code                  : chr [1:1117] "94806" "94109-8160" "93906" "94131" ...
##  $ pat_race                  : chr [1:1117] "Black or African American" "Black or African American"
##  $ ethnicity                 : chr [1:1117] "Not Hispanic or Latino" "Not Hispanic or Latino" "Hispa
##  $ marital_status            : chr [1:1117] "Single" "Single" "Married" "Single" ...
##  $ language                  : chr [1:1117] "English" "English" "Spanish" "English" ...
##  $ insurance_type            : chr [1:1117] "01 - Medicare" "01 - Medicare" "03 - Private Coverage"
##  $ smoking                   : chr [1:1117] "Former Smoker" "Never Smoker" "Never Smoker" "Former Sm
##  $ BMI                       : num [1:1117] 24.1 30.9 31.1 35 28.3 ...
##  $ covid_tested              : chr [1:1117] "Yes" "Yes" "Yes" "Yes" ...
##  $ covid_pos                 : chr [1:1117] "Yes" "Yes" "Yes" "Yes" ...
##  $ covid_pos_time            : POSIXct[1:1117], format: "2020-07-17 19:11:00" "2020-07-31 10:41:00"
##  $ RR_encounter              : chr [1:1117] "Yes" "No" "No" "No" ...
##  $ RR_event_time             : chr [1:1117] "2020-04-26 14:57:00.000" "NULL" "NULL" "NULL" ...
##  $ NIV_encounter             : chr [1:1117] "Yes" "No" "No" "No" ...
##  $ NIV_event_time            : chr [1:1117] "2020-04-26 12:15:00.000" "NULL" "NULL" "NULL" ...
##  $ new_intubation_encounter  : chr [1:1117] "Yes" "No" "No" "No" ...
##  $ new_intubation_event_time : chr [1:1117] "2020-04-26 15:40:00.000" "NULL" "NULL" "NULL" ...
##  $ admit_or_transfer_ICU     : chr [1:1117] "Yes" "No" "No" "No" ...
##  $ admit_or_transfer_ICU_time: chr [1:1117] "2020-04-26 11:59:00.000" "NULL" "NULL" "NULL" ...
##  $ ECMO_encounter            : chr [1:1117] "No" "No" "No" "No" ...
##  $ ECMO_start_time           : chr [1:1117] "NULL" "NULL" "NULL" "NULL" ...
##  $ HD_UF_CRRT_encounter      : chr [1:1117] "Yes" "No" "No" "No" ...
##  $ HD_UF_CRRT_start_time     : chr [1:1117] "2020-04-22 20:00:00.000" "NULL" "NULL" "NULL" ...
##  $ end_in_death              : chr [1:1117] "No" "No" "No" "No" ...
##  $ death_time                : chr [1:1117] "NULL" "NULL" "NULL" "NULL" ...
##  $ readmit                   : chr [1:1117] "Yes" "No" "No" "No" ...
##  $ HAI_encounter             : chr [1:1117] "No" "No" "No" "No" ...
##  $ HAI_type                  : chr [1:1117] "NULL" "NULL" "NULL" "NULL" ...
##  $ HAI_pathogen              : chr [1:1117] "NULL" "NULL" "NULL" "NULL" ...
##  $ HAI_first_date            : chr [1:1117] "NULL" "NULL" "NULL" "NULL" ...
##  $ facility_name             : chr [1:1117] "PARN - UCSF MEDICAL CENTER HOSPITAL - 505 PARNASSUS AVE
##  $ admit_hospital_service    : chr [1:1117] "Advanced Heart Failure" "Hospital Medicine" "Hospital M
##  $ enc_class                 : chr [1:1117] "Inpatient" "Inpatient" "Inpatient" "Inpatient" ...
##  $ encounter_start_time      : POSIXct[1:1117], format: "2020-04-22 11:10:00" "2020-07-31 06:31:00"
##  $ ED_disp_time              : chr [1:1117] "2020-04-22 12:09:00.000" "2020-07-31 17:13:00.000" "202
##  $ leave_ed_time             : chr [1:1117] "2020-04-22 14:34:00.000" "2020-07-31 21:30:00.000" "202
##  $ ED_dispo                  : chr [1:1117] "Admit" "Admit" "Admit" "Admit" ...
##  $ admit_time                : POSIXct[1:1117], format: "2020-04-22 12:09:00" "2020-07-31 16:03:00"
##  $ admit_diag_text           : chr [1:1117] "Sepsis (HCC)" "CHEST PAIN" "flank pain, concern for cur
##  $ admit_floor_unit          : chr [1:1117] "10LS CVT" "MZ 4 EAST" "15L ADULT ACUTE CARE" "15L ADULT
##  $ present_source            : chr [1:1117] "Self Referred (Home)" "Self Referred (Home)" "Self Refe
##  $ arrival_means             : chr [1:1117] "Assist From Vehicle" "Ground Ambulance" "Walk In" "Walk
##  $ arrival_escort_by         : chr [1:1117] "Self" "Self" "Family Member" "Self" ...
##  $ discharge_time            : POSIXct[1:1117], format: "2020-05-29 14:28:00" "2020-08-06 13:07:00"
##  $ discharge_disposition     : chr [1:1117] "Home Health Care (Non UCSF)" "Home or Self Care" "Home
```

```
##  $ hospital_LOS              : num [1:1117] 37 6 2 3 4 9 1 6 25 2 ...
##  $ DRG_NUMBER                : chr [1:1117] "MS870" "MS177" "APR720" "MS393" ...
##  $ DRG_NAME                  : chr [1:1117] "SEPTICEMIA OR SEVERE SEPSIS WITH MV >96 HOURS" "RESPIRA
##  - attr(*, "spec")=
##   .. cols(
##   ..   deid_enc_id = col_character(),
##   ..   deid_pat_id = col_character(),
##   ..   age = col_double(),
##   ..   sex = col_character(),
##   ..   zip_code = col_character(),
##   ..   pat_race = col_character(),
##   ..   ethnicity = col_character(),
##   ..   marital_status = col_character(),
##   ..   language = col_character(),
##   ..   insurance_type = col_character(),
##   ..   smoking = col_character(),
##   ..   BMI = col_double(),
##   ..   covid_tested = col_character(),
##   ..   covid_pos = col_character(),
##   ..   covid_pos_time = col_datetime(format = ""),
##   ..   RR_encounter = col_character(),
##   ..   RR_event_time = col_character(),
##   ..   NIV_encounter = col_character(),
##   ..   NIV_event_time = col_character(),
##   ..   new_intubation_encounter = col_character(),
##   ..   new_intubation_event_time = col_character(),
##   ..   admit_or_transfer_ICU = col_character(),
##   ..   admit_or_transfer_ICU_time = col_character(),
##   ..   ECMO_encounter = col_character(),
##   ..   ECMO_start_time = col_character(),
##   ..   HD_UF_CRRT_encounter = col_character(),
##   ..   HD_UF_CRRT_start_time = col_character(),
##   ..   end_in_death = col_character(),
##   ..   death_time = col_character(),
##   ..   readmit = col_character(),
##   ..   HAI_encounter = col_character(),
##   ..   HAI_type = col_character(),
##   ..   HAI_pathogen = col_character(),
##   ..   HAI_first_date = col_character(),
##   ..   facility_name = col_character(),
##   ..   admit_hospital_service = col_character(),
##   ..   enc_class = col_character(),
##   ..   encounter_start_time = col_datetime(format = ""),
##   ..   ED_disp_time = col_character(),
##   ..   leave_ed_time = col_character(),
##   ..   ED_dispo = col_character(),
##   ..   admit_time = col_datetime(format = ""),
##   ..   admit_diag_text = col_character(),
##   ..   admit_floor_unit = col_character(),
##   ..   present_source = col_character(),
##   ..   arrival_means = col_character(),
##   ..   arrival_escort_by = col_character(),
##   ..   discharge_time = col_datetime(format = ""),
##   ..   discharge_disposition = col_character(),
```

```
##   ..   hospital_LOS = col_double(),
##   ..   DRG_NUMBER = col_character(),
##   ..   DRG_NAME = col_character()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

Create vector of variables to include in final dataset, assign to `vars`

```
vars <- c("deid_enc_id", "age", "sex", "zip_code", "pat_race", "ethnicity", "smoking",
          "BMI", "admit_or_transfer_ICU", "end_in_death", "death_time", "admit_time",
          "discharge_time", "hospital_LOS")
```

Clean and transform `dm` to expanded dataset where each row presents patient-days (each row is one patient per day of hospitalization)

```
dm <- dm %>%
  filter(covid_pos == "Yes" & hospital_LOS > 0) %>%     # Filter dataset where `covid_pos` == 'Yes' & `
  select(vars) %>%            # Filter to only include variables in `var` vector
  rename(ID = deid_enc_id,
         race = pat_race,
         zip = zip_code,
         LOS = hospital_LOS) %>%      # Parse date from `admit_time` and `discharge_time`
  mutate(admit_date = as.Date(admit_time),
         discharge_date = as.Date(discharge_time),
         death_time = as.POSIXct(death_time, format = "%Y-%m-%d %H:%M:%S"),
         death_date = as.Date(death_time)) %>%
  group_by(ID) %>%  # Group by 'ID' (unique patient identifier)
  mutate(day_date = list(seq(min(admit_date), max(discharge_date), by = "day"))) %>%
  unnest(day_date)    # Create new var `day_date` that spans from `admit_date` to `discharge_date` for
```

```
## Note: Using an external vector in selections is ambiguous.
## i Use ‘all_of(vars)‘ instead of ‘vars‘ to silence this message.
## i See <https://tidyselect.r-lib.org/reference/faq-external-vector.html>.
## This message is displayed once per session.
```

Transform and collapse factor variables

```
# Transform `sex` to factor
dm$sex <- as.factor(dm$sex)

# Transform `race` to factor, and collapse categories
dm$race <- as.factor(dm$race)
```

```
dm$race <- fct_collapse(dm$race,
                        "Other" = c("Other", "American Indian or Alaska Native"),
                        "Unknown" = c("Unknown", "Declined", "Unknown/Declined"),
                        "Native Hawaiian or Other Pacific Islander" = c("Native Hawaiian or Other Pacif

# Transform `ethnicity` to factor, and collapse categories
dm$ethnicity <- as.factor(dm$ethnicity)

dm$ethnicity <- fct_collapse(dm$ethnicity,
                             "Unknown" = c("Declined", "Unknown", "Unknown/Declined"))

# Transform `smoking` to factor, and collapse categories
dm$smoking <- as.factor(dm$smoking)

dm$smoking <- fct_collapse(dm$smoking,
                           "Current Smoker" = c("Current Every Day Smoker", "Current Some Day Smoker",
                                                "Smoker, Current Status Unknown"),
                           "Not Current Smoker" = c("Never Smoker", "Former Smoker", "Passive Smoke Exp
                           "Smoking Status Unknown" = c("Unknown If Ever Smoked", "NULL"))
```

Save final `dm` dataset as .Rdata and export as .csv

```
write_csv(dm, here("data", "dm_covid_11.08.21.csv"))
save(dm, file = here("data", "dm_covid.Rdata"))
```

End of Document