dataset36_withmed_withEKG

Fay Yan

2024-07-22

load data dataset: newDatabase_36_ANON

```
library(readxl)
newDatabase_36_ANON <- read_excel("~/Desktop/newDatabase_36_ANON.xlsx",
    sheet = "Asymptomatic_CAC", skip = 2)
```

```
## New names:
## * `Date` -> `Date...135`
## * `Date` -> `Date...137`
## * `Date` -> `Date...139`
## * `Date` -> `Date...141`
## * `Excision` -> `Excision...149`
## * `Excision Date` -> `Excision Date...150`
## * `Radiation` -> `Radiation...151`
## * `Radiation Date` -> `Radiation Date...152`
## * `Chemo` -> `Chemo...153`
## * `Chemo Date` -> `Chemo Date...154`
## * `Excision` -> `Excision...158`
## * `Excision Date` -> `Excision Date...159`
## * `Radiation` -> `Radiation...160`
## * `Radiation Date` -> `Radiation Date...161`
## * `Chemo` -> `Chemo...162`
## * `Chemo Date` -> `Chemo Date...163`
## * `Excision` -> `Excision...166`
## * `Excision Date` -> `Excision Date...167`
## * `Radiation` -> `Radiation...168`
## * `Radiation Date` -> `Radiation Date...169`
## * `Chemo` -> `Chemo...170`
## * `Chemo Date` -> `Chemo Date...171`
## * `Chemo Date` -> `Chemo Date...179`
```

select column: (including variables for medication and EKG)

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

ANON36_selected_EKG <- newDatabase_36_ANON[, c(1:6,11,40,41,43,44,45,47,49,
                                               50:57,61:64,65:66,75:86,92:95,
                                               96,107:115)]
```

For risk factors, encode to binary (change NA to 0)

```
ANON36_selected_EKG <- ANON36_selected_EKG %>%
  mutate(across(15:22, ~ ifelse(is.na(.) | . == "", 0, .)))
```

add column Male:

```
ANON36_selected_EKG <- ANON36_selected_EKG %>%
  mutate(Male = ifelse(Gender == "Male", 1, 0))
```

handle Race: "American Indian or Alaskan Native" = 1 "Asian" = 2 "Black" = 3 "Hispanic" = 4 "Other" = 5 "Unknown" = 6 "White" = 7 NA = missing

```
# Step 1: Convert "NA" strings to actual NA values
ANON36_selected_EKG$Race[ANON36_selected_EKG$Race == "NA"] <- NA
# Step 2: Convert Race to a factor
ANON36_selected_EKG$Race <- as.factor(ANON36_selected_EKG$Race)
# Step 3: Manually encode Race as numeric, keeping NA as NA
ANON36_selected_EKG$Race_encoded <- ifelse(is.na(ANON36_selected_EKG$Race), NA, as.numeric(ANON36_select
```

For medication columns, encode to binary (change NA to 0)

```
ANON36_EKG_withmed <- ANON36_selected_EKG %>%
  mutate(across(c(`Beta Blockers`, `Antianginal Agents`, `Antihyperlipidemic`, `Calcium Blockers`, `Ant
                  `Antidiabetic`, `Antiarrhythmic`),
              ~ ifelse(grepl("prior|on date", ., ignore.case = TRUE), 1, 0)))
```

EKG columns: select time period (within 1 year) to keep EKG column values

```
# Convert the columns to Date format
ANON36_EKG_withmed <- ANON36_EKG_withmed %>%
  mutate(`Exam Completion Timestamp` = as.Date(`Exam Completion Timestamp`, format = "%Y-%m-%d"),
         `EKG Date` = as.Date(`EKG Date`, format = "%Y-%m-%d"))

# Check if the dates are within one year
ANON36_EKG_withmed <- ANON36_EKG_withmed %>%
  mutate(within_one_year = abs(difftime(`Exam Completion Timestamp`, `EKG Date`, units = "days")) <= 365

ANON36_EKG <- ANON36_EKG_withmed %>%
  mutate(across(30:44, ~ case_when(
    within_one_year ~ .,
    TRUE ~ NA
  )))
```

MACE dataset: when either MACE or Death = 1, result = 1; otherwise 0

```r
ANON36_MACE_EKG <- ANON36_EKG[, -c(2,3,5,6,27:29,57)]
ANON36_MACE_EKG <- ANON36_MACE_EKG %>%
  mutate(result = if_else(!is.na(`First MACE or Death (1=MACE, 2=Death)`), 1, 0))
ANON36_MACE_EKG_output <- ANON36_MACE_EKG[, -c(38)]


write.csv(ANON36_MACE_EKG_output, "/Users/Student/Desktop/ANON36_MACE_EKG_output.csv", row.names = FALSE
```

HF dataset: when CHF date after exam date, result = 1; otherwise 0

```r
ANON36_HF_EKG <- ANON36_EKG[, -c(3,5,6,29,45,57)]
ANON36_HF_EKG$`Exam Completion Timestamp` <- as.Date(ANON36_HF_EKG$`Exam Completion Timestamp`, format =
ANON36_HF_EKG$`CHF Dx Date` <- as.Date(ANON36_HF_EKG$`CHF Dx Date`,format = "%Y-%m-%d")
ANON36_HF_EKG <- ANON36_HF_EKG %>%
  mutate(is_later = case_when(
    !is.na(`Exam Completion Timestamp`) & !is.na(`CHF Dx Date`) ~ `CHF Dx Date` > `Exam Completion Times
    TRUE ~ NA
  ))
ANON36_HF_EKG <- ANON36_HF_EKG %>%
  mutate(result = if_else(is.na(is_later) | is_later == FALSE, 0, 1))
ANON36_HF_EKG_output <- ANON36_HF_EKG[, -c(2,24,25,52)]


write.csv(ANON36_HF_EKG_output, "/Users/Student/Desktop/ANON36_HF_EKG_output.csv", row.names = FALSE)
```

death dataset: when Death = 1, result = 1; otherwise 0

```r
ANON36_death_EKG <- ANON36_EKG[, -c(2,3,5,6,27:29,57)]
ANON36_death_EKG <- ANON36_death_EKG %>%
  mutate(`First MACE or Death (1=MACE, 2=Death)` = ifelse(is.na(`First MACE or Death (1=MACE, 2=Death)`
  mutate(result = ifelse(`First MACE or Death (1=MACE, 2=Death)` == 2, 1, 0))
ANON36_death_EKG_output <- ANON36_death_EKG[, -c(38)]


write.csv(ANON36_death_EKG_output, "/Users/Student/Desktop/ANON36_death_EKG_output.csv", row.names = FA
```