Palatnik eda

Maggie Westerland

2023-02-28

```
df <- read.csv('~/Desktop/capstone/ob_gyn_mother_baby.csv')
```

data dictionary:

patient_num: *integer* Masked MRN, the i2b2 patient number

mom_admit_dttm_shifted: *date* date of the mother's delivery admission

mom_disch_ddtm_shifted: *date* date of the mother's delivery discharge

mom_los: *integer* mom's length of stay in days for the delivery encounter

mom_age_at_del: *integer* mom's age at delivery in years

mom_marital_status: *factor 4 levels – Divorced, Married, Single, Other* mom's current marital status

mom_race: *factor 3 levels – White or Caucasian, Black or African American, Other* mom's current race

mom_ethnicity: *factor 3 levels – Hispanic, Non-Hispanic, Other* mom's current ethnicity

mom_ht_ft: *numeric* mom's height (feet counterpart)

mom_ht_in: *numeric* mom's height (inches counterpart)

mom_wt_oz: *integer* mom's weight in ounces

mom_bmi: *numeric* mom's body mass index

preeclampsia: *factor 2 levels – Yes, No* indicates presence of a preeclampsia diagnosis on the delivery encounter

pregest_dm: *factor 2 levels – Yes, No* indicates presence of a pregestational diabetes diagnosis on the delivery encounter

gestestational_dm: *factor 2 levels – Yes, No* indicates presence of a gestational diabetes diagnosis on the delivery encounter

placental_abruption: *factor 2 levels – Yes, No* indicates presence of a placental abruption diabetes diagnosis on the delivery encounter

maternal_dvt: *factor 2 levels – Yes, No* indicates presence of a maternal deep vein thrombosis diabetes diagnosis on the delivery encounter

baby_patient_num: *integer* Masked MRN, the i2b2 patient number

baby_yob_shifted: *integer* baby's year of birth

baby_dob_shifted: *date* baby's date of birth

baby_dod_shifted: *date* baby's date of death

baby_admit_dttm_shifted: *date* date of the baby's birth admission

baby__disch__dttm__shifted: *date* date of the baby's birth discharge

baby__birth__ht__in: *numeric* baby's birth height in inches

baby__birth__wt__oz: *numeric* baby's birth weight in ounces

delivery method: *factor 4 levels – C-Section, Operative Vaginal, Vaginal, Other* name of the delivery method used for the birth

baby__sex: *factor 2 levels – Male, Female* baby's current sex

baby__race: *factor 3 levels – White or Caucasian, Black or African American, Other* baby's current race

baby__ethnicity: *factor 3 levels – Hispanic, Non-Hispanic, Other* baby's current ethnicity

```
df$mom_admit_dttm_shifted <-
  gsub(" .*", "", df$mom_admit_dttm_shifted)
df$mom_disch_dttm_shifted <-
  gsub(" .*", "", df$mom_disch_dttm_shifted)
df$baby_dob_shifted <- gsub(" .*", "", df$baby_dob_shifted)
df$baby_dod_shifted <- gsub(" .*", "", df$baby_dod_shifted)
df$baby_admit_dttm_shifted <-
  gsub(" .*", "", df$baby_admit_dttm_shifted)
df$baby_disch_dttm_shifted <-
  gsub(" .*", "", df$baby_disch_dttm_shifted)

df_mutated <- df %>%
  add_column(mom_ht_ft = NA) %>%
  add_column(mom_ht_in = NA) %>%
  select(
    -delivery_encounter_num,
    -del_dept_id,
    -mom_patient_num,
    -del_dept_name,
    -mom_disch_disp,
    -del_stage1_hrs,
    -del_stage2_hrs,
    -del_stage3_hrs,
    -bloodloss_del,
    -bloodloss_total,
    -birth_encounter_num,
    -baby_del_dept,
    -baby_los,
    -baby_disch_disp,
    -gest_age_wks,
    -apgar1,
    -apgar5,
    -apgar10,
    -anesth_conc,
    -del_meth_id,
    -ob_gravidity,
    -ob_parity,
    -ob_multiple_births
  ) %>%
  mutate(
    mom_ethnicity = case_when(
      mom_ethnicity == 'Refused' ~ 'Other',
```

```
    mom_ethnicity == 'Hispanic' ~ 'Hispanic',
    mom_ethnicity == 'Unknown' ~ 'Other',
    mom_ethnicity == 'NI' ~ 'Other',
    mom_ethnicity == 'Non Hispanic' ~ 'NH'
  )
) %>%
mutate(
  mom_race = case_when(
    mom_race == 'Unknown' ~ 'Other',
    mom_race == 'Asian' ~ 'Other',
    mom_race == 'Patient Refused' ~ 'Other',
    mom_race == 'Multiracial' ~ 'Other',
    mom_race == 'Native Hawaiian or Other Pacific Islander' ~ 'Other',
    mom_race == 'American Indian or Alaska Native' ~ 'Other',
    mom_race == 'White or Caucasian' ~ 'White or Caucasian',
    mom_race == 'Black or African American' ~ 'Black or African American',
    mom_race == 'Other' ~ 'Other',
  )
) %>%
mutate(
  baby_ethnicity = case_when(
    baby_ethnicity == 'Refused' ~ 'Other',
    baby_ethnicity == 'Unknown' ~ 'Other',
    baby_ethnicity == 'Hispanic' ~ 'Hispanic',
    baby_ethnicity == 'NI' ~ 'Other',
    baby_ethnicity == 'Non Hispanic' ~ 'NH'
  )
) %>%
mutate(
  baby_race = case_when(
    baby_race == 'Unknown' ~ 'Other',
    baby_race == 'Asian' ~ 'Other',
    baby_race == 'Patient Refused' ~ 'Other',
    baby_race == 'Multiracial' ~ 'Other',
    baby_race == 'Native Hawaiian or Other Pacific Islander' ~ 'Other',
    baby_race == 'American Indian or Alaska Native' ~ 'Other',
    baby_race == 'White or Caucasian' ~ 'White or Caucasian',
    baby_race == 'Black or African American' ~ 'Black or African American',
    baby_race == 'Other' ~ 'Other',
  )
) %>%
mutate(
  mom_marital_status = case_when(
    mom_marital_status == 'Legally Separated' ~ 'Divorced',
    mom_marital_status == 'Divorced' ~ 'Divorced',
    mom_marital_status == 'Patient Refused' ~ 'Other',
    mom_marital_status == 'Unknown' ~ 'Other',
    mom_marital_status == 'Widowed' ~ 'Other',
    mom_marital_status == 'Significant Other' ~ 'Other',
    mom_marital_status == 'Married' ~ 'Married',
    mom_marital_status == 'Single' ~ 'Single'
  )
) %>%
```

```
mutate(mom_marital_status = as.factor(mom_marital_status)) %>%
mutate(mom_race = as.factor(mom_race)) %>%
mutate(mom_ethnicity = as.factor(mom_ethnicity)) %>%
mutate(preeclampsia = as.factor(preeclampsia)) %>%
mutate(pregest_dm = as.factor(pregest_dm)) %>%
mutate(gestational_dm = as.factor(gestational_dm)) %>%
mutate(placental_abruption = as.factor(placental_abruption)) %>%
mutate(maternal_dvt = as.factor(maternal_dvt)) %>%
mutate(baby_birth_wt_gms = as.numeric(baby_birth_wt_gms)) %>%
mutate(
  delivery_method = case_when(
    delivery_method == 'C-Section' ~ 'C-Section',
    delivery_method == 'C-Section w/ BTL' ~ 'C-Section',
    delivery_method == 'C-Section, Classical' ~ 'C-Section',
    delivery_method == 'C-Section, Hysterectomy' ~ 'C-Section',
    delivery_method == 'C-Section, Low Transverse' ~ 'C-Section',
    delivery_method == 'C-Section, Low Vertical' ~ 'C-Section',
    delivery_method == 'C-Section, Unspecified' ~ 'C-Section',
    delivery_method == 'Elective Abortion' ~ 'Other',
    delivery_method == 'Forceps Vaginal Delivery' ~ 'Operative Vaginal',
    delivery_method == 'Miscarriage' ~ 'Other',
    delivery_method == 'NST' ~ 'Other',
    delivery_method == 'Other/Procedure' ~ 'Other',
    delivery_method == 'Spontaneous Abortion' ~ 'Other',
    delivery_method == 'TAB' ~ 'Other',
    delivery_method == 'Vacuum Vaginal Delivery' ~ 'Operative Vaginal',
    delivery_method == 'Vaginal, Breech' ~ 'Vaginal',
    delivery_method == 'Vaginal, Breech Extraction' ~ 'Vaginal',
    delivery_method == 'Vaginal, Forceps' ~ 'Operative Vaginal',
    delivery_method == 'Vaginal, Forceps Delivery' ~ 'Operative Vaginal',
    delivery_method == 'Vaginal, Spontaneous' ~ 'Vaginal',
    delivery_method == 'Vaginal, Spontaneous Breech' ~ 'Vaginal',
    delivery_method == 'Vaginal, Vacuum (Extractor)' ~ 'Operative Vaginal',
    delivery_method == 'Vaginal, Vacuum Delivery' ~ 'Operative Vaginal',
    delivery_method == 'VBAC, Forceps Delivery' ~ 'Operative Vaginal',
    delivery_method == 'VBAC, Spontaneous' ~ 'Vaginal',
    delivery_method == 'VBAC, Vacuum Delivery' ~ 'Operative Vaginal',
  )
) %>%
mutate(delivery_method = as.factor(delivery_method)) %>%
mutate(
  baby_sex = case_when(
    baby_sex == 'Unknown' ~ '',
    baby_sex == 'Male' ~ 'Male',
    baby_sex == 'Female' ~ 'Female'
  )
) %>%
separate(mom_ht_ftin,
         into = c('mom_ht_ft', 'mom_ht_in'),
         sep = "'") %>%
mutate(baby_sex = as.factor(baby_sex)) %>%
mutate(baby_race = as.factor(baby_race)) %>%
mutate(baby_ethnicity = as.factor(baby_ethnicity)) %>%
```

```
  mutate(baby_admit_dttm_shifted = as.Date(baby_admit_dttm_shifted, "%m/%d/%y")) %>%
  mutate(baby_disch_dttm_shifted = as.Date(baby_disch_dttm_shifted, "%m/%d/%y")) %>%
  mutate(baby_dob_shifted = as.Date(baby_dob_shifted, "%m/%d/%y")) %>%
  mutate(baby_dod_shifted = as.Date(baby_dod_shifted, "%m/%d/%y")) %>%
  mutate(mom_admit_dttm_shifted = as.Date(mom_admit_dttm_shifted, "%m/%d/%y")) %>%
  mutate(mom_disch_dttm_shifted = as.Date(mom_disch_dttm_shifted, "%m/%d/%y"))

df_mutated$mom_ht_in <- gsub('\"' , "", df_mutated$mom_ht_in)

df_mutated <- df_mutated %>%
  mutate(mom_ht_ft = as.numeric(mom_ht_ft)) %>%
  mutate(mom_ht_in = as.numeric(mom_ht_in))

df_mutated[df_mutated == ""] <- NA
```

```
str(df_mutated, vec.len = 2, strict.width = 'cut')
```

```
## 'data.frame':    32499 obs. of  30 variables:
##  $ patient_num            : int  29732601 19678083 32945117 23004068 28381336 ..
##  $ mom_admit_dttm_shifted : Date, format: "2016-10-30" "2019-11-12" ...
##  $ mom_disch_dttm_shifted : Date, format: "2016-11-02" "2019-11-14" ...
##  $ mom_los                : int  3 2 2 2 2 ...
##  $ mom_age_at_del         : int  34 42 18 33 24 ...
##  $ mom_marital_status     : Factor w/ 4 levels "Divorced","Married",..: 2 2 4 ..
##  $ mom_race               : Factor w/ 3 levels "Black or African American",..:..
##  $ mom_ethnicity          : Factor w/ 3 levels "Hispanic","NH",..: 2 2 1 2 2 ...
##  $ mom_ht_ft              : num  5 5 5 5 5 ...
##  $ mom_ht_in              : num  0 1 3 2 7 ...
##  $ mom_wt_oz              : int  3616 2832 2144 2608 2864 ...
##  $ mom_bmi                : num  44.1 33.4 ...
##  $ preeclampsia           : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 ...
##  $ pregest_dm             : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 ...
##  $ gestational_dm         : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 ...
##  $ placental_abruption    : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 ...
##  $ maternal_dvt           : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 ...
##  $ baby_patient_num       : int  28775049 29913097 26379596 18846057 41430988 ..
##  $ baby_yob_shifted       : int  2016 2019 2020 2020 2018 ...
##  $ baby_dob_shifted       : Date, format: "2016-10-30" "2019-11-12" ...
##  $ baby_dod_shifted       : Date, format: NA NA ...
##  $ baby_admit_dttm_shifted: Date, format: "2016-10-30" "2019-11-12" ...
##  $ baby_disch_dttm_shifted: Date, format: "2016-11-02" "2019-11-14" ...
##  $ baby_birth_ht_in       : num  19 19 20 19 21 ...
##  $ baby_birth_wt_oz       : num  132 128 ...
##  $ baby_birth_wt_gms      : num  3730 3640 ...
##  $ delivery_method        : Factor w/ 4 levels "C-Section","Operative Vaginal"..
##  $ baby_sex               : Factor w/ 3 levels "","Female","Male": 2 2 2 2 3 ...
##  $ baby_race              : Factor w/ 3 levels "Black or African American",..:..
##  $ baby_ethnicity         : Factor w/ 3 levels "Hispanic","NH",..: 2 2 1 2 2 ...
```

```
save(df_mutated, file="~/Desktop/capstone/Palatnik_2_22.Rdata")
```