EHR Data Quality Control and Analysis

Fereshteh Izadi

2025-07-19

```r
# Set path to data directory
data_path <- "C:/Users/User/Downloads/test_data/data/dest/"

# Load main datasets
patients <- read_csv(paste0(data_path, "patients.csv"))
```

```
## Rows: 1162 Columns: 25
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (19): Id, BIRTHDATE, SSN, DRIVERS, PASSPORT, PREFIX, FIRST, LAST, SUFFI...
## dbl   (5): ZIP, LAT, LON, HEALTHCARE_EXPENSES, HEALTHCARE_COVERAGE
## date  (1): DEATHDATE
##
## 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.
```

```r
conditions <- read_csv(paste0(data_path, "conditions.csv"))
```

```
## Rows: 38100 Columns: 5
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (2): PATIENT, ENCOUNTER
## dbl  (1): CODE
## date (2): START, STOP
##
## 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.
```

```r
observations <- read_csv(paste0(data_path, "observations.csv"))
```

```
## Rows: 531144 Columns: 8
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (7): PATIENT, ENCOUNTER, CATEGORY, CODE, VALUE, UNITS, TYPE
## dttm (1): DATE
##
## 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.
```

```r
medications <- read_csv(paste0(data_path, "medications.csv"))
```

```
## Warning: One or more parsing issues, call `problems()` on your data frame for details,
## e.g.:
##   dat <- vroom(...)
##   problems(dat)
```

```
## Rows: 56430 Columns: 11
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (3): PATIENT, PAYER, ENCOUNTER
## dbl  (6): CODE, BASE_COST, PAYER_COVERAGE, DISPENSES, TOTALCOST, REASONCODE
## dttm (2): START, STOP
##
## 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.
```

```r
encounters <- read_csv(paste0(data_path, "encounters.csv"))
```

```
## Rows: 61459 Columns: 13
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (6): Id, PATIENT, ORGANIZATION, PROVIDER, PAYER, ENCOUNTERCLASS
## dbl  (5): CODE, BASE_ENCOUNTER_COST, TOTAL_CLAIM_COST, PAYER_COVERAGE, REASO...
## dttm (2): START, STOP
##
## 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.
```

```r
# Load dictionaries
dict_snomed <- read_csv(paste0(data_path, "dictionary_snomed.csv"))
```

```
## Rows: 204 Columns: 2
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): DESCRIPTION
## dbl (1): CODE
##
## 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.
```

```r
dict_rxnorm <- read_csv(paste0(data_path, "dictionary_rxnorm.csv"))
```

```
## Rows: 176 Columns: 2
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): DESCRIPTION
## dbl (1): CODE
##
## 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.
```

```
dict_loinc  <- read_csv(paste0(data_path, "dictionary_loinc.csv"))
```

```
## Rows: 256 Columns: 2
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (2): CODE, DESCRIPTION
##
## 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.
```

Patients Quality Control (QC)

```
# Convert date columns
patients$BIRTHDATE <- as.Date(patients$BIRTHDATE)
patients$DEATHDATE <- as.Date(patients$DEATHDATE)

# Calculate age
patients$AGE <- as.numeric(floor(interval(patients$BIRTHDATE, Sys.Date()) / years(1)))

# Filter out invalid data
patients <- patients %>%
  filter(AGE >= 0 & AGE <= 100) %>%
  filter(is.na(DEATHDATE) | BIRTHDATE <= DEATHDATE) %>%
  filter(!(LAT == 0 & LON == 0)) %>%
  filter(LAT >= -90 & LAT <= 90) %>%
  filter(LON >= -180 & LON <= 180) %>%
  filter(!(HEALTHCARE_EXPENSES == 0 & HEALTHCARE_COVERAGE > 5000)) %>%
  filter(GENDER %in% c("M", "F", "", NA)) %>%
  filter(!((PREFIX == "Mr." & GENDER == "F") |
           (PREFIX %in% c("Mrs.", "Ms.", "Miss") & GENDER == "M")))

# Remove US state + UK city mismatches
us_states <- state.abb
uk_cities <- c("London", "Manchester", "Leeds", "Birmingham", "Glasgow", "Liverpool")
patients <- patients[!(patients$CITY %in% uk_cities & patients$STATE %in% us_states), ]

# Fix race coding
patients$RACE[patients$RACE == "XJniDSe"] <- "other (possibly miscoded)"

# ZIP formatting
patients$ZIP <- sprintf("%05d", patients$ZIP)
```

Encounters QC

```
encounters$START <- as.POSIXct(encounters$START)
encounters$STOP <- as.POSIXct(encounters$STOP)

encounters <- encounters %>%
  filter(is.na(STOP) | STOP >= START) %>%
```

```
  filter(BASE_ENCOUNTER_COST >= 0, TOTAL_CLAIM_COST >= 0, PAYER_COVERAGE >= 0) %>%
  filter(PATIENT %in% patients$Id)
```

Conditions QC

```
conditions$START <- as.Date(conditions$START)
conditions$STOP <- as.Date(conditions$STOP)

conditions <- conditions %>%
  filter(is.na(STOP) | START <= STOP) %>%
  filter(PATIENT %in% patients$Id) %>%
  filter(ENCOUNTER %in% encounters$Id) %>%
  filter(CODE %in% dict_snomed$CODE)

conditions <- left_join(conditions, dict_snomed, by = "CODE")
```

```
## Warning in left_join(conditions, dict_snomed, by = "CODE"): Detected an unexpected many-to-many rela
## i Row 304 of `x` matches multiple rows in `y`.
## i Row 15 of `y` matches multiple rows in `x`.
## i If a many-to-many relationship is expected, set `relationship =
##   "many-to-many"` to silence this warning.
```

```
# Add 0/1 flags for socioeconomic factors based on DESCRIPTION content
conditions <- conditions %>%
  mutate(
    factor_education = as.integer(grepl("education|high school|primary school|higher education", DESCRI
    factor_employment = as.integer(grepl("employment|unemployed|labor force", DESCRIPTION, ignore.case =
    factor_stress = as.integer(grepl("stress|anxiety|panic", DESCRIPTION, ignore.case = TRUE)),
    factor_refugee = as.integer(grepl("refugee", DESCRIPTION, ignore.case = TRUE)),
    factor_criminal = as.integer(grepl("criminal", DESCRIPTION, ignore.case = TRUE)),
    factor_violence = as.integer(grepl("violence|abuse|victim", DESCRIPTION, ignore.case = TRUE)),
    factor_transport = as.integer(grepl("transport", DESCRIPTION, ignore.case = TRUE)),
    factor_drugs_alcohol = as.integer(grepl("alcohol|misuses drugs", DESCRIPTION, ignore.case = TRUE)),
    factor_social_isolation = as.integer(grepl("social|isolation|limited contact", DESCRIPTION, ignore.
    factor_military = as.integer(grepl("armed forces|military", DESCRIPTION, ignore.case = TRUE)),
    factor_housing = as.integer(grepl("housing", DESCRIPTION, ignore.case = TRUE))
  )
```

Observations QC

```
observations$DATE <- as.POSIXct(observations$DATE)
observations$CODE <- as.character(observations$CODE)
dict_loinc$CODE <- as.character(dict_loinc$CODE)

observations <- observations %>%
  left_join(dict_loinc, by = "CODE") %>%
  filter(!is.na(DESCRIPTION), !is.na(VALUE), VALUE != "", !is.na(TYPE), TYPE != "")
```

```
## Warning in left_join(., dict_loinc, by = "CODE"): Detected an unexpected many-to-many relationship be
## i Row 10 of `x` matches multiple rows in `y`.
## i Row 1 of `y` matches multiple rows in `x`.
## i If a many-to-many relationship is expected, set `relationship =
##   "many-to-many"` to silence this warning.
```

```
observations$VALUE_NUMERIC <- suppressWarnings(as.numeric(observations$VALUE))
observations <- observations[observations$TYPE != "numeric" | !is.na(observations$VALUE_NUMERIC), ]

# Weight-kg mismatch
mismatch <- grepl("weight", tolower(observations$DESCRIPTION)) & !grepl("kg", tolower(observations$UNITS
observations <- observations[!mismatch, ]

observations <- observations %>%
  filter(PATIENT %in% patients$Id, ENCOUNTER %in% encounters$Id)
```

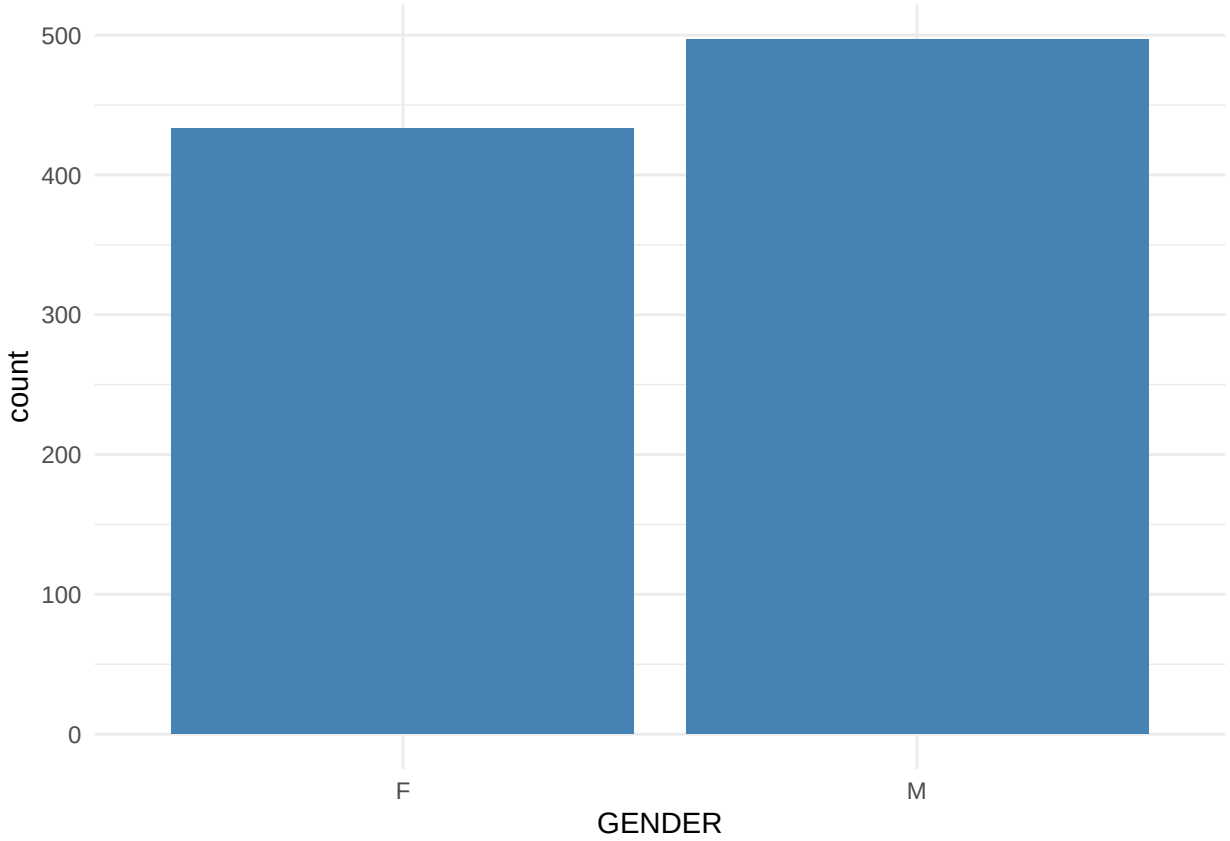
Medications QC

```
medications$START <- as.POSIXct(medications$START)
medications$STOP <- as.POSIXct(medications$STOP)

medications <- medications %>%
  filter(is.na(STOP) | START <= STOP, PATIENT %in% patients$Id, ENCOUNTER %in% encounters$Id) %>%
  filter(BASE_COST >= 0, PAYER_COVERAGE >= 0, TOTALCOST >= 0)

medications <- medications %>%
  left_join(dict_rxnorm, by = "CODE") %>%
  filter(!is.na(DESCRIPTION)) %>%
  filter(BASE_COST <= TOTALCOST) %>%
  filter(abs((BASE_COST + PAYER_COVERAGE) - TOTALCOST) <= 1)
```

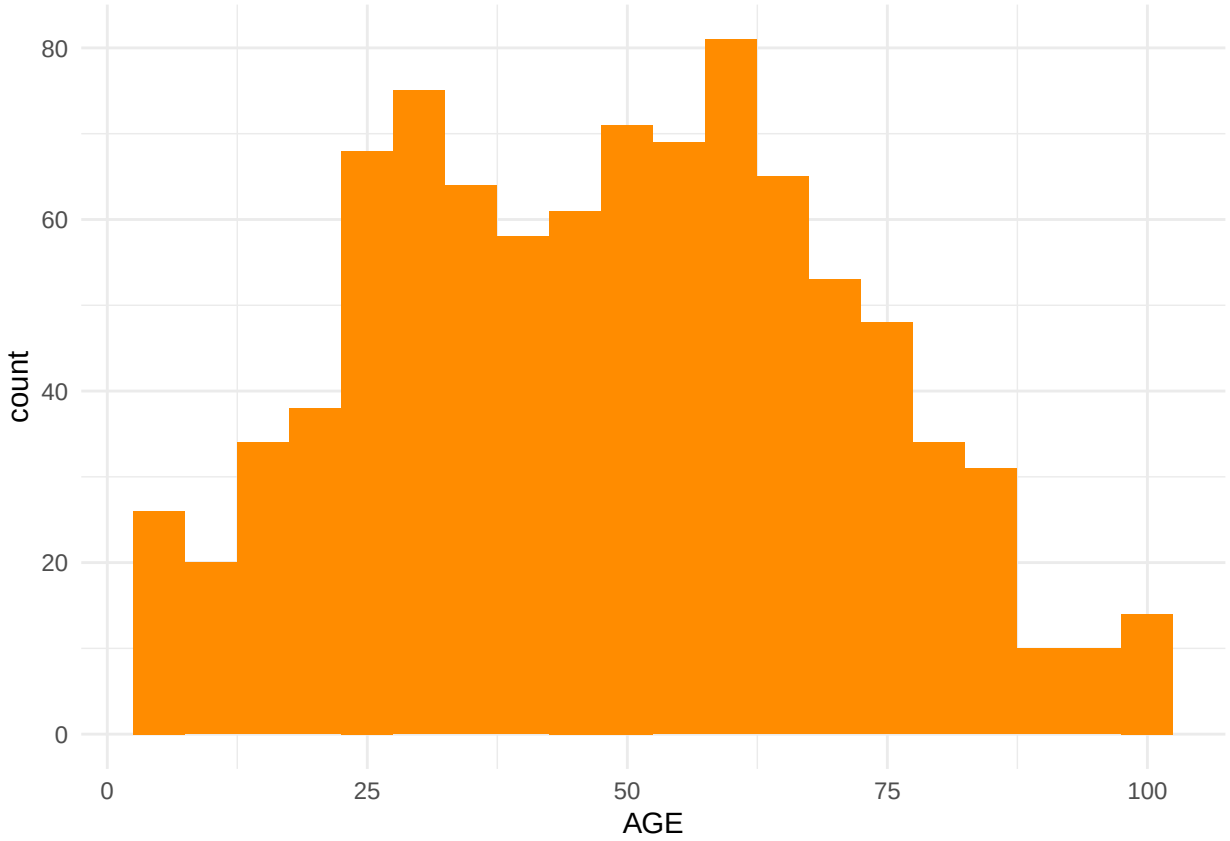
```
## Warning in left_join(., dict_rxnorm, by = "CODE"): Detected an unexpected many-to-many relationship b
## i Row 51 of `x` matches multiple rows in `y`.
## i Row 9 of `y` matches multiple rows in `x`.
## i If a many-to-many relationship is expected, set `relationship =
##   "many-to-many"` to silence this warning.
```

Descriptive Summaries and Visualizations

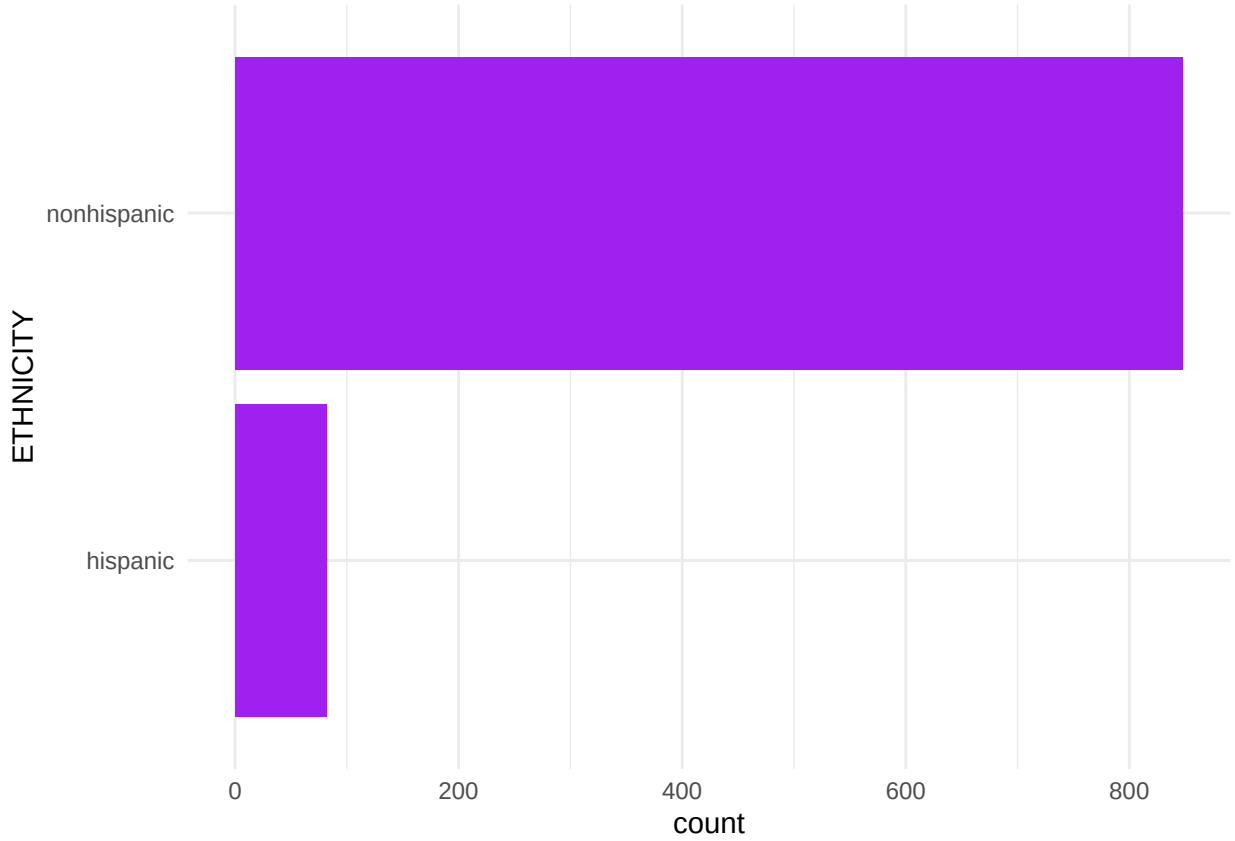
```
# Gender distribution
ggplot(patients, aes(x = GENDER)) + geom_bar(fill = "steelblue") + theme_minimal()
```



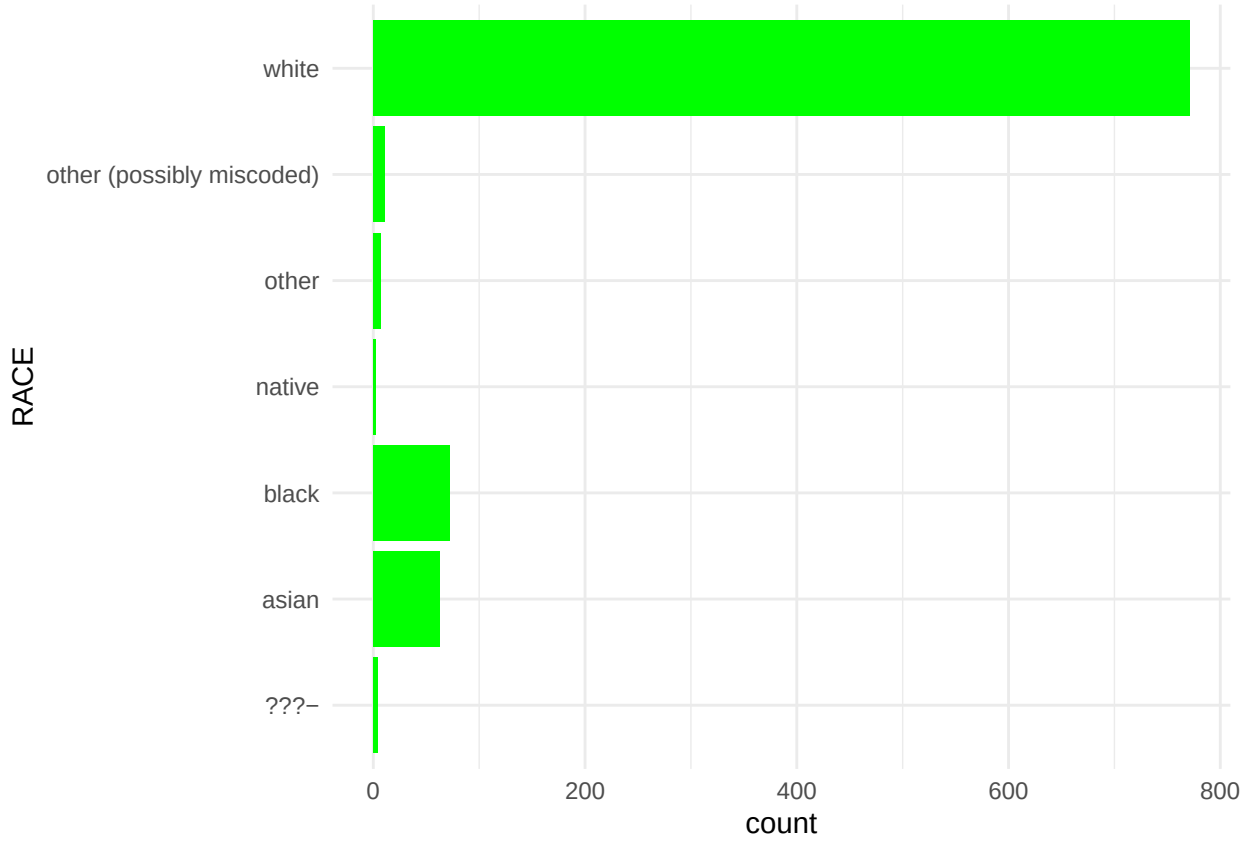

```
# Age distribution
ggplot(patients, aes(x = AGE)) + geom_histogram(binwidth = 5, fill = "darkorange") + theme_minimal()
```




```r
# Ethnicity and Race
ggplot(patients, aes(x = ETHNICITY)) + geom_bar(fill = "purple") + coord_flip() + theme_minimal()
```




```
ggplot(patients, aes(x = RACE)) + geom_bar(fill = "green") + coord_flip() + theme_minimal()
```



Top Medications, Conditions, and Labs

```
top_meds <- medications %>%
  group_by(PATIENT, DESCRIPTION) %>% summarise(n = n()) %>%
  group_by(DESCRIPTION) %>% summarise(total = sum(n)) %>% arrange(desc(total)) %>% head(10)
```

```
## `summarise()` has grouped output by 'PATIENT'. You can override using the
## `.groups` argument.
```

```
conditions_summary <- conditions %>%
  group_by(DESCRIPTION) %>% summarise(n = n()) %>% arrange(desc(n))
top_conditions <- head(conditions_summary, 10)
least_conditions <- tail(conditions_summary, 10)

top_labs <- observations %>%
  group_by(DESCRIPTION) %>% summarise(n = n()) %>% arrange(desc(n)) %>% head(10)
```

Conditions with Highest Out-of-Pocket Costs

```
cond_cost <- merge(conditions, encounters[, c("Id", "TOTAL_CLAIM_COST", "PAYER_COVERAGE")],
                   by.x = "ENCOUNTER", by.y = "Id") %>%
```

```
  group_by(DESCRIPTION) %>%
  summarise(
    avg_claim_cost = mean(TOTAL_CLAIM_COST, na.rm = TRUE),
    avg_patient_pay = mean(TOTAL_CLAIM_COST - PAYER_COVERAGE, na.rm = TRUE)) %>%
  arrange(desc(avg_patient_pay))
```

Reasons for Visit

```
reasons <- encounters %>%
  filter(!is.na(REASONCODE)) %>%
  count(REASONCODE, sort = TRUE) %>%
  left_join(dict_snomed, by = c("REASONCODE" = "CODE")) %>%
  rename(ReasonDescription = DESCRIPTION)
head(reasons, 20)
```

```
## # A tibble: 20 x 3
##    REASONCODE     n ReasonDescription
##         <dbl> <int> <chr>
##  1   72892002  3585 Normal pregnancy
##  2   55822004  1640 Hyperlipidemia
##  3   88805009  1536 Chronic congestive heart failure (disorder)
##  4  444814009  1210 Viral sinusitis (disorder)
##  5   10509002   633 Acute bronchitis (disorder)
##  6  195662009   632 Acute viral pharyngitis (disorder)
##  7  254837009   380 Malignant neoplasm of breast (disorder)
##  8  271737000   284 Anemia (disorder)
##  9  192127007   280 Child attention deficit disorder
## 10   59621000   242 Hypertension
## 11   75498004   213 Acute bacterial sinusitis (disorder)
## 12  195967001   193 Asthma
## 13   36971009   176 Sinusitis (disorder)
## 14   55680006   145 Drug overdose
## 15   43878008   131 Streptococcal sore throat (disorder)
## 16   65363002   122 Otitis media
## 17  233678006   115 Childhood asthma
## 18   74400008   106 Appendicitis
## 19   82423001    99 Chronic pain
## 20  196416002    89 Impacted molars
```

Done

```
summary(patients)
```

```
##       Id              BIRTHDATE            DEATHDATE
##  Length:930         Min.   :1924-12-04   Min.   :1945-10-20
##  Class :character   1st Qu.:1960-03-30   1st Qu.:1997-01-05
##  Mode  :character   Median :1975-11-20   Median :2008-07-02
##                     Mean   :1976-06-17   Mean   :2005-04-22
```

```
##                   3rd Qu.:1994-07-28   3rd Qu.:2015-05-22
##                   Max.   :2021-09-05   Max.   :2021-09-29
##                                        NA's   :833
##      SSN              DRIVERS            PASSPORT            PREFIX
##  Length:930         Length:930         Length:930         Length:930
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##     FIRST               LAST              SUFFIX             MAIDEN
##  Length:930         Length:930         Length:930         Length:930
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##    MARITAL              RACE             ETHNICITY            GENDER
##  Length:930         Length:930         Length:930         Length:930
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##   BIRTHPLACE          ADDRESS              CITY              STATE
##  Length:930         Length:930         Length:930         Length:930
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##     COUNTY              ZIP                 LAT              LON
##  Length:930         Length:930         Min.   :41.50   Min.   :-73.45
##  Class :character   Class :character   1st Qu.:42.10   1st Qu.:-71.43
##  Mode  :character   Mode  :character   Median :42.31   Median :-71.13
##                                        Mean   :42.25   Mean   :-71.30
##                                        3rd Qu.:42.45   3rd Qu.:-70.99
##                                        Max.   :42.89   Max.   :-69.98
##
##  HEALTHCARE_EXPENSES HEALTHCARE_COVERAGE      AGE
##  Min.   :     826    Min.   :      0     Min.   :  3.00
##  1st Qu.:  589426    1st Qu.:   1675     1st Qu.: 30.25
##  Median : 1248438    Median :  22745     Median : 49.00
##  Mean   : 1305014    Mean   : 158385     Mean   : 48.58
##  3rd Qu.: 1914155    3rd Qu.: 214946     3rd Qu.: 65.00
##  Max.   :12641789    Max.   :2411214     Max.   :100.00
##
```

```
summary(encounters)
```

```
##       Id                 START                         STOP
##  Length:49339       Min.   :1925-10-01 21:07:04   Min.   :1925-10-01 21:22:04
##  Class :character   1st Qu.:2001-10-31 20:33:41   1st Qu.:2001-11-01 06:28:47
##  Mode  :character   Median :2013-07-16 01:52:59   Median :2013-07-16 02:35:10
##                     Mean   :2007-08-27 23:09:14   Mean   :2007-08-28 04:39:02
##                     3rd Qu.:2017-09-04 05:48:47   3rd Qu.:2017-09-04 06:03:47
##                     Max.   :2021-11-19 16:50:22   Max.   :2021-11-19 17:05:22
##
##    PATIENT           ORGANIZATION         PROVIDER            PAYER
##  Length:49339       Length:49339       Length:49339       Length:49339
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##  ENCOUNTERCLASS          CODE           BASE_ENCOUNTER_COST TOTAL_CLAIM_COST
##  Length:49339       Min.   :  1505002   Min.   : 77.49      Min.   :     0.0
##  Class :character   1st Qu.:162673000   1st Qu.: 77.49      1st Qu.:   129.2
##  Mode  :character   Median :185347001   Median :129.16      Median :   786.3
##                     Mean   :260883600   Mean   :113.68      Mean   :  3976.6
##                     3rd Qu.:390906007   3rd Qu.:129.16      3rd Qu.:  1615.7
##                     Max.   :702927004   Max.   :129.16      Max.   :873646.2
##
##  PAYER_COVERAGE        REASONCODE
##  Min.   :     0.00   Min.   :1.734e+06
##  1st Qu.:     0.00   1st Qu.:6.256e+07
##  Median :     0.00   Median :7.289e+07
##  Mean   :   866.70   Mean   :5.840e+12
##  3rd Qu.:    20.14   3rd Qu.:1.957e+08
##  Max.   :227851.81   Max.   :1.094e+16
##                      NA's   :36206
```

```
summary(medications)
```

```
##      START                         STOP                          PATIENT
##  Min.   :1931-03-09 03:45:19   Min.   :1931-03-27 03:45:19   Length:17008
##  1st Qu.:1997-10-31 14:27:06   1st Qu.:1998-01-23 06:20:21   Class :character
##  Median :2008-11-30 14:43:20   Median :2009-01-07 18:44:36   Mode  :character
##  Mean   :2005-06-22 07:29:37   Mean   :2005-07-25 16:08:31
##  3rd Qu.:2016-06-09 03:36:53   3rd Qu.:2016-06-16 16:47:30
##  Max.   :2021-11-18 14:01:22   Max.   :2021-11-18 14:01:22
##                                NA's   :512
##    PAYER            ENCOUNTER             CODE           BASE_COST
##  Length:17008       Length:17008       Min.   : 106258   Min.   :   0.01
##  Class :character   Class :character   1st Qu.: 310798   1st Qu.:   0.01
##  Mode  :character   Mode  :character   Median : 310798   Median :   0.02
##                                        Mean   : 428470   Mean   : 383.43
##                                        3rd Qu.: 314231   3rd Qu.:  60.57
##                                        Max.   :2123111   Max.   :6994.54
##
##  PAYER_COVERAGE        DISPENSES          TOTALCOST         REASONCODE
##  Min.   :0.0000000   Min.   :    1.00   Min.   :   0.01   Min.   : 10509002
##  1st Qu.:0.0000000   1st Qu.:    1.00   1st Qu.:   0.12   1st Qu.: 55822004
```

```
##  Median :0.0000000   Median :    1.00   Median :   0.56   Median : 59621000
##  Mean   :0.0001529   Mean   :   12.06   Mean   : 383.52   Mean   :111924994
##  3rd Qu.:0.0000000   3rd Qu.:   12.00   3rd Qu.:  60.57   3rd Qu.: 59621000
##  Max.   :0.6200000   Max.   :45000.00   Max.   :6994.54   Max.   :706870000
##                                                           NA's   :2880
##  DESCRIPTION
##  Length:17008
##  Class :character
##  Mode  :character
##
##
##
##
```

```
summary(conditions)
```

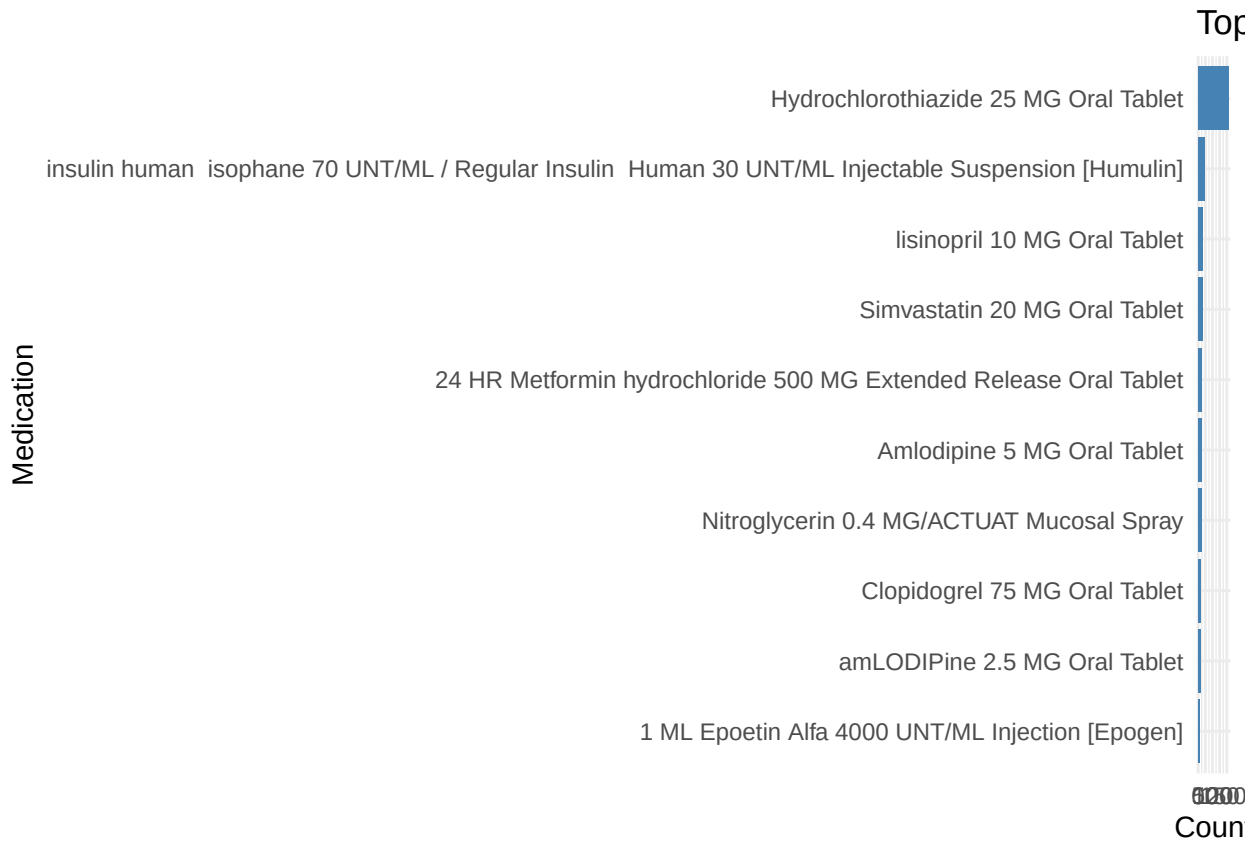
```
##      START                STOP              PATIENT
##  Min.   :1930-08-05   Min.   :1930-10-21   Length:31334
##  1st Qu.:1996-04-30   1st Qu.:2000-05-09   Class :character
##  Median :2011-07-23   Median :2013-06-20   Mode  :character
##  Mean   :2005-01-29   Mean   :2007-07-07
##  3rd Qu.:2017-02-08   3rd Qu.:2018-03-07
##  Max.   :2021-11-15   Max.   :2021-11-18
##                       NA's   :7014
##   ENCOUNTER             CODE           DESCRIPTION        factor_education
##  Length:31334       Min.   :1.734e+06   Length:31334       Min.   :0.00000
##  Class :character   1st Qu.:9.130e+07   Class :character   1st Qu.:0.00000
##  Mode  :character   Median :1.609e+08   Mode  :character   Median :0.00000
##                     Mean   :5.463e+13                      Mean   :0.02582
##                     3rd Qu.:2.376e+08                      3rd Qu.:0.00000
##                     Max.   :1.094e+16                      Max.   :1.00000
##
##  factor_employment factor_stress    factor_refugee   factor_criminal
##  Min.   :0.0000    Min.   :0.0000   Min.   :0.0000   Min.   :0.000000
##  1st Qu.:0.0000    1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.000000
##  Median :0.0000    Median :0.0000   Median :0.0000   Median :0.000000
##  Mean   :0.4488    Mean   :0.1356   Mean   :0.0015   Mean   :0.005234
##  3rd Qu.:1.0000    3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.000000
##  Max.   :1.0000    Max.   :1.0000   Max.   :1.0000   Max.   :1.000000
##
##  factor_violence   factor_transport   factor_drugs_alcohol
##  Min.   :0.00000   Min.   :0.000000   Min.   :0.000000
##  1st Qu.:0.00000   1st Qu.:0.000000   1st Qu.:0.000000
##  Median :0.00000   Median :0.000000   Median :0.000000
##  Mean   :0.03763   Mean   :0.006159   Mean   :0.007245
##  3rd Qu.:0.00000   3rd Qu.:0.000000   3rd Qu.:0.000000
##  Max.   :1.00000   Max.   :1.000000   Max.   :1.000000
##
##  factor_social_isolation factor_military    factor_housing
##  Min.   :0.00000         Min.   :0.000000   Min.   :0.000000
##  1st Qu.:0.00000         1st Qu.:0.000000   1st Qu.:0.000000
##  Median :0.00000         Median :0.000000   Median :0.000000
##  Mean   :0.06692         Mean   :0.001213   Mean   :0.004564
##  3rd Qu.:0.00000         3rd Qu.:0.000000   3rd Qu.:0.000000
```

```
##  Max.   :1.00000         Max.   :1.000000   Max.   :1.000000
##
```

```r
summary(observations)
```

```
##       DATE                          PATIENT           ENCOUNTER
##  Min.   :1935-11-30 10:04:43   Length:439021      Length:439021
##  1st Qu.:2013-07-17 12:00:54   Class :character   Class :character
##  Median :2016-08-01 05:23:50   Mode  :character   Mode  :character
##  Mean   :2015-01-28 21:03:31
##  3rd Qu.:2019-07-24 23:14:15
##  Max.   :2021-11-18 16:26:22
##
##    CATEGORY             CODE              VALUE              UNITS
##  Length:439021      Length:439021      Length:439021      Length:439021
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##      TYPE           DESCRIPTION        VALUE_NUMERIC
##  Length:439021      Length:439021      Min.   :    -3.7
##  Class :character   Class :character   1st Qu.:     5.3
##  Mode  :character   Mode  :character   Median :    29.2
##                                        Mean   :  3083.0
##                                        3rd Qu.:    92.0
##                                        Max.   :957744.0
##                                        NA's   :168873
```

Medications by Socioeconomic Group



Note that the `echo = TRUE` parameter is added to the code chunk to prevent printing of the R code that generated the plot.