Lab1-Q2&3

2023-01-17

```
data <- read.csv("icd10cm_codes_2020.txt",header = F, sep = "\t" )
#load data into a dataframe
library(stringr)
data_Q2 <- str_split_fixed(data$V1, " ", 2)
colnames(data_Q2) <- c("ICD10", "description")
#split columns and rename
data_Q2 <- data.frame(data_Q2)
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
```

```
data_Q2 %>% select(contains(c("A", "B")))
```

```
## data frame with 0 columns and 72184 rows
```

```
#select A and B
```

```
data_Q3 <- read.csv("DE1_0_2008_to_2010_Inpatient_Claims_Sample_1.csv",header = T)
#load data into a dataframe
data_Q3 <- data_Q3[order(data_Q3$CLM_FROM_DT),]
#sort the data in time order with CLM_FROM_DT
data_Q3 <- distinct(data_Q3, DESYNPUF_ID, .keep_all = TRUE)
#delete the repetitive IDs, keep the first-appeared data, so the rest is first admiss
ion
```

```
library(dplyr)
library(tidyselect)
#select data with opioid abuse ralated codes and place the IDs in a new table
cnt <- 0
table <- as.data.frame(matrix(nrow=37780,ncol=2))
colnames(table) <- c("ID", "Race")
for(i in 1:37780){
  for (j in 21:30){
    if(data_Q3[i,j] == 30500 | data_Q3[i,j] == 30551 | data_Q3[i,j] == 30552 | data_Q
3[i,j] == 30553)
    {cnt <- cnt + 1
    table$ID[cnt] <- data_Q3$DESYNPUF_ID[i]
    break
    }
  }
}
len_table <- cnt
table <- data.frame(table[1:len_table,])
```

```
data_info <- read.csv("DE1_0_2008_Beneficiary_Summary_File_Sample_1.csv",header = T)
#load the data and calculate the length
len_info <- length(data_info$DESYNPUF_ID)
#select the opioid abuse IDs from data and place the corresponding race into the tabl
e
for(i in 1:len_table){
  for(j in 1:len_info){
    if(table$ID[i] == data_info$DESYNPUF_ID[j]){
      table$Race[i] <- data_info$BENE_RACE_CD[j]
      break
    }
  }
}
```

```
library(ggplot2)
library(psych)
```

```
##
## Attaching package: 'psych'
```

```
## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha
```

```
race_info <- data.frame(table(table$Race))
race_info
```

```
##   Var1 Freq
## 1    1  166
## 2    2   31
## 3    3    6
## 4    5    6
```

```
#use table to get descriptions of race, put into a new table, and draw a barplot reco
rdingly
colnames(race_info) <- c("Race", "Freq")
barplot(race_info$Freq,  main="Race Distribution of Opioid Overuse", col=c("#6666CC",
"#3366FF","#009999","#00CC00"), names.arg = c("White","Black","Other","Hispanic"))
```