Lab1

Guanyu Zou

2023-01-13

Question 2

Question 2a

Using this file create a data frame that has two columns: “ICD10” and “description”.

```
library(tidyr)
library(tidyverse)
```

```
## ── Attaching packages ─────────────────────────────────────── tidyve
rse 1.3.2 ──
## ✓ ggplot2 3.3.6      ✓ dplyr   1.0.10
## ✓ tibble  3.1.8      ✓ stringr 1.4.1
## ✓ readr   2.1.2      ✓ forcats 0.5.2
## ✓ purrr   0.3.4
## ── Conflicts ────────────────────────────────────────── tidyverse_co
nflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
```

```
icd <- read.table(file = "~/Desktop/icd10cm_codes_2020.txt", sep = "\t")
```

```
## Warning in scan(file = file, what = what, sep = sep, quote = quote,
dec = dec, :
## EOF within quoted string
```

```
icd <- separate(icd, V1, into = c("ICD10", "description"), sep = "^\\S*
\\K\\s+")

head(icd, 5)
```

```
##   ICD10                                          description
## 1  A000 Cholera due to Vibrio cholerae 01, biovar cholerae
## 2  A001    Cholera due to Vibrio cholerae 01, biovar eltor
## 3  A009                                Cholera, unspecified
## 4 A0100                          Typhoid fever, unspecified
## 5 A0101                                  Typhoid meningitis
```

Question 2b

From the created data frame find a number of different diagnoses for the first chapter “Certain infectious and parasitic diseases” with codes start at “A00” and end at “B99”.

```
icd_ch1A <- icd %>% filter(str_detect(ICD10, "^A"))
icd_ch1B <- icd %>% filter(str_detect(ICD10, "^B"))
nrow(icd_ch1A) + nrow(icd_ch1B)

## [1] 902
```

The total number of different diagnoses for the first chapter “Certain infectious and parasitic diseases” is 902.

```
head(icd, 10)

##    ICD10                                          description
## 1   A000 Cholera due to Vibrio cholerae 01, biovar cholerae
## 2   A001    Cholera due to Vibrio cholerae 01, biovar eltor
## 3   A009                                 Cholera, unspecified
## 4  A0100                           Typhoid fever, unspecified
## 5  A0101                                 Typhoid meningitis
## 6  A0102               Typhoid fever with heart involvement
## 7  A0103                                  Typhoid pneumonia
## 8  A0104                                  Typhoid arthritis
## 9  A0105                               Typhoid osteomyelitis
## 10 A0109             Typhoid fever with other complications
```

Above output is the first 10 diagnoses for the first chapter “Certain infectious and parasitic diseases”.

Question 3

Question 3a

Select only first admission for each patient.

```
df1 <- read.csv("~/Desktop/DE1_0_2008_to_2010_Inpatient_Claims_Sample_1.
csv")
df1a <- df1 %>% group_by(DESYNPUF_ID) %>% filter(CLM_ADMSN_DT == min(CL
M_ADMSN_DT))
head(df1a, c(5, 5))

## # A tibble: 5 × 5
## # Groups:   DESYNPUF_ID [5]
##   DESYNPUF_ID       CLM_ID SEGMENT CLM_FROM_DT CLM_THRU_DT
##   <chr>              <dbl>   <int>       <int>       <int>
## 1 00013D2EFD8E45D1 1.97e14       1    20100312    20100313
## 2 00016F745862898F 1.96e14       1    20090412    20090418
## 3 00052705243EA128 1.97e14       1    20080912    20080912
```

```
## 4 0007F12A492FD25D 1.97e14        1     20080919     20080922
## 5 000B97BA2314E971 1.96e14        1     20091209     20091213
```

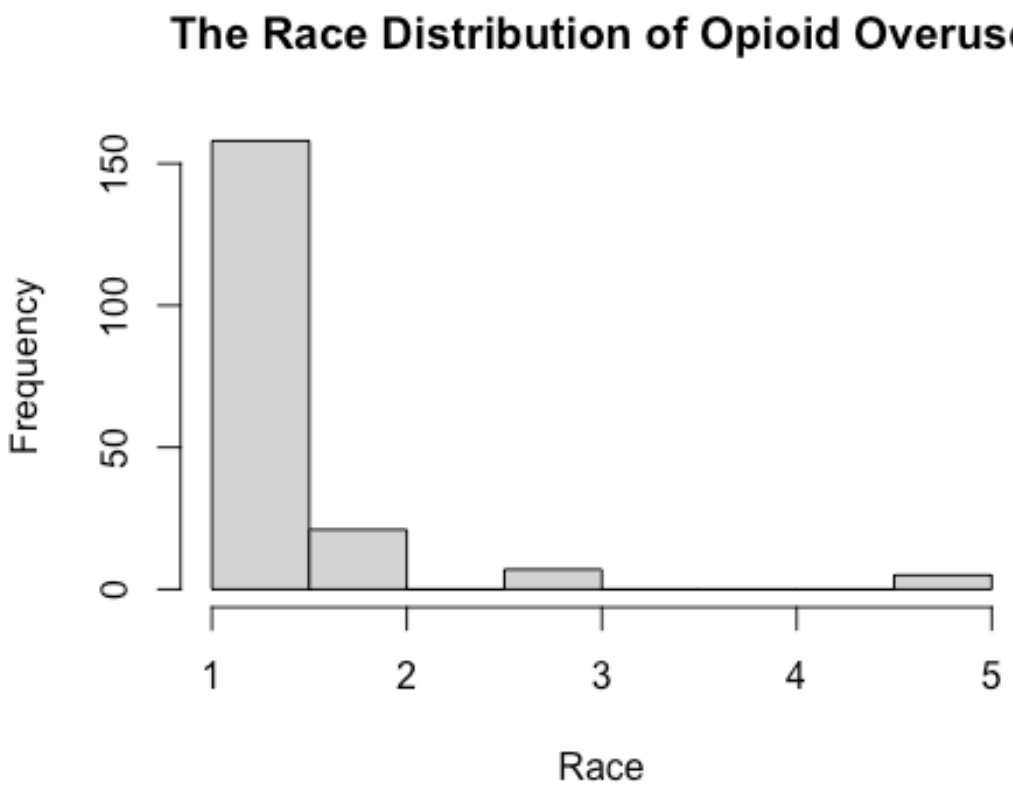
Question 3b

Using both files, find the race distribution of opioid overuse.

```
df2 <- read.csv("~/Desktop/DE1_0_2008_Beneficiary_Summary_File_Sample_1.
csv")
df1b <- df1a %>% filter(CLM_DRG_CD == "304" | CLM_DRG_CD == "305")
tableq3 <- df2 %>% left_join(df1b) %>% drop_na(CLM_DRG_CD)

## Joining, by = "DESYNPUF_ID"

hist(tableq3$BENE_RACE_CD, main = "The Race Distribution of Opioid Over
use",
     xlab = "Race")
```



Question 3c

Comment on your results.

First, we know the code meaning are: 1:WHITE; 2:BLACK; 3:OTHER; 5:HISPANIC. As we can see in graph, the distribution are extreme right skewed which means there

are a lot of white people and less black/hispanic/other, and no asian. Whites accounted for almost all opioid overuse patients.