MIS 500 Porfolio

Michael Pearson

1/27/2020

This is some stuff to get writing going

```
## Connect to the page and download the zipped files

temp <- tempfile()

download.file("https://archive.ics.uci.edu/ml/machine-learning-databases/00296/dataset_diabetes.zip", te

post_zipped <- unzip(temp)

print(post_zipped)

## [1] "./dataset_diabetes/diabetic_data.csv"
## [2] "./dataset_diabetes/IDs_mapping.csv"

diabetes_data <- read.csv("./dataset_diabetes/diabetic_data.csv", header = TRUE)

print(head(diabetes_data))

##   encounter_id patient_nbr            race gender     age weight
## 1      2278392     8222157       Caucasian Female  [0-10)      ?
## 2       149190    55629189       Caucasian Female [10-20)      ?
## 3        64410    86047875 AfricanAmerican Female [20-30)      ?
## 4       500364    82442376       Caucasian   Male [30-40)      ?
## 5        16680    42519267       Caucasian   Male [40-50)      ?
## 6        35754    82637451       Caucasian   Male [50-60)      ?
##   admission_type_id discharge_disposition_id admission_source_id
## 1                 6                       25                   1
## 2                 1                        1                   7
## 3                 1                        1                   7
## 4                 1                        1                   7
## 5                 1                        1                   7
## 6                 2                        1                   2
##   time_in_hospital payer_code      medical_specialty num_lab_procedures
## 1                1          ? Pediatrics-Endocrinology                 41
## 2                3          ?                        ?                 59
## 3                2          ?                        ?                 11
## 4                2          ?                        ?                 44
## 5                1          ?                        ?                 51
## 6                3          ?                        ?                 31
##   num_procedures num_medications number_outpatient number_emergency
## 1              0               1                 0                0
## 2              0              18                 0                0
## 3              5              13                 2                0
## 4              1              16                 0                0
## 5              0               8                 0                0
## 6              6              16                 0                0
##   number_inpatient diag_1 diag_2 diag_3 number_diagnoses max_glu_serum
```

```
## 1                  0 250.83      ?      ?                  1           None
## 2                  0   276 250.01    255                  9           None
## 3                  1   648    250    V27                  6           None
## 4                  0     8 250.43    403                  7           None
## 5                  0   197    157    250                  5           None
## 6                  0   414    411    250                  9           None
##   A1Cresult metformin repaglinide nateglinide chlorpropamide glimepiride
## 1      None        No          No          No             No          No
## 2      None        No          No          No             No          No
## 3      None        No          No          No             No          No
## 4      None        No          No          No             No          No
## 5      None        No          No          No             No          No
## 6      None        No          No          No             No          No
##   acetohexamide glipizide glyburide tolbutamide pioglitazone rosiglitazone
## 1            No        No        No          No           No            No
## 2            No        No        No          No           No            No
## 3            No    Steady        No          No           No            No
## 4            No        No        No          No           No            No
## 5            No    Steady        No          No           No            No
## 6            No        No        No          No           No            No
##   acarbose miglitol troglitazone tolazamide examide citoglipton insulin
## 1       No       No           No         No      No          No      No
## 2       No       No           No         No      No          No      Up
## 3       No       No           No         No      No          No      No
## 4       No       No           No         No      No          No      Up
## 5       No       No           No         No      No          No  Steady
## 6       No       No           No         No      No          No  Steady
##   glyburide.metformin glipizide.metformin glimepiride.pioglitazone
## 1                  No                  No                       No
## 2                  No                  No                       No
## 3                  No                  No                       No
## 4                  No                  No                       No
## 5                  No                  No                       No
## 6                  No                  No                       No
##   metformin.rosiglitazone metformin.pioglitazone change diabetesMed
## 1                      No                     No     No           No
## 2                      No                     No     Ch          Yes
## 3                      No                     No     No          Yes
## 4                      No                     No     Ch          Yes
## 5                      No                     No     Ch          Yes
## 6                      No                     No     No          Yes
##   readmitted
## 1         NO
## 2        >30
## 3         NO
## 4         NO
## 5         NO
## 6        >30
```

```
## Convert to a data table for easier subsetting

## Now convert to a data table so that the genders can be broken out into separate tables for analysis

library(data.table)
```

```r
diabetes_data <- data.table(diabetes_data)
setkey(diabetes_data, gender)
males <- diabetes_data["Male"]
females <- diabetes_data["Female"]
setkey(diabetes_data, readmitted)
nope <- diabetes_data["NO"]
more_than_30 <- diabetes_data[">30"]
less_than_30 <- diabetes_data["<30"]

setkey(nope, gender)
male_nope <- nope["Male"]
female_nope <- nope["Female"]
setkey(more_than_30, gender)
male_more_than_30 <- more_than_30["Male"]
female_more_than_30 <- more_than_30["Female"]
setkey(less_than_30, gender)
male_less_than_30 <- less_than_30["Male"]
female_less_than_30 <- less_than_30["Female"]


## Or another way

gender_test <- table(diabetes_data$gender, diabetes_data$readmitted)

all_data_gender <- chisq.test(gender_test)

## Warning in chisq.test(gender_test): Chi-squared approximation may be
## incorrect
print(all_data_gender)

##
##  Pearson's Chi-squared test
##
## data:  gender_test
## X-squared = 37.461, df = 4, p-value = 1.447e-07
race_test <- table(diabetes_data$race, diabetes_data$readmitted)

all_data_race <- chisq.test(race_test)

print(all_data_race)

##
##  Pearson's Chi-squared test
##
## data:  race_test
## X-squared = 282.59, df = 10, p-value < 2.2e-16
age_test <- table(diabetes_data$age, diabetes_data$readmitted)

all_data_age <- chisq.test(age_test)

print(all_data_age)

##
```

```
##  Pearson's Chi-squared test
##
## data:  age_test
## X-squared = 313.17, df = 18, p-value < 2.2e-16
```

```r
## Removing patients with multiple admissions

diabetes_data$duplicates <- duplicated(diabetes_data$patient_nbr)

Dupe_diabetes_data <- diabetes_data[diabetes_data$duplicates == "TRUE"]

dupe_gender_table <- table(Dupe_diabetes_data$gender, Dupe_diabetes_data$readmitted)

dupes_gender_test <- chisq.test(dupe_gender_table)
```

```
## Warning in chisq.test(dupe_gender_table): Chi-squared approximation may be
## incorrect
```

```r
print(dupes_gender_test)
```

```
##
##  Pearson's Chi-squared test
##
## data:  dupe_gender_table
## X-squared = NaN, df = 4, p-value = NA
```

```r
## Patients numbers only once (may have been readmitted)



No_duplicated_data <- diabetes_data[diabetes_data$duplicates =="FALSE"]


## Patient numbers used more than once


Back_again <- diabetes_data[diabetes_data$duplicates =="TRUE"]
```