Data 607 Final Project

Warner Alexis

2023-12-10

Hospital Readmission Reduction

The center for Medicare and Medicaid Services begun to reduce payment to Hospitals for excessive readmissions on October 1rst 2012 as part of the Affordable Care Act. Hospitals' mission switches strategies to reduce rehospitalization rate and improves quality care so patients don't come back within 30 days readmission. There are several strategies implemented to enable the process but, the use of data analytics has been indispensable to reduction of readmission rate. Warchol et al. said: "Data analytics can be used to improve clinical operations, watch for care patterns, and identify readmission risk." He acknowledges that other researcher like Monga suggested that hospitals have the ability to design an analytical model to predict the likelihood of patients' readmission on the basis of information collected in Electronic Health Records (EHR). The purpose of this project is to predict the hospital readmission from this data set in UC Irvine Machine Learning Repository called "Diabetes 130-US hospitals for years 1999-2008".

This data set contains information about care given to patients in 130 Hospitals from 1990- 2008. It has 50 columns representing patients and hospital outcomes.

Data Source: https://archive.ics.uci.edu/dataset/296/diabetes+130-us+hospitals+for+years+1999-2008 (https://archive.ics.uci.edu/dataset/296/diabetes+130-us+hospitals+for+years+1999-2008)

Readmission is considered when patient return to hospital 30 days after his first admission.

── Data Summary ──────────────────────────── Values
Name diabetic Number of rows 101766
Number of columns 52

```
library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✓ dplyr     1.1.3     ✓ readr     2.1.4
## ✓ forcats   1.0.0     ✓ stringr   1.5.0
## ✓ ggplot2   3.4.3     ✓ tibble    3.2.1
## ✓ lubridate 1.9.2     ✓ tidyr     1.3.0
## ✓ purrr     1.0.2
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(ggplot2)
library(skimr)
library(htmlTable)
#library(glmnet)
library(caret)
```

```
## Loading required package: lattice
##
## Attaching package: 'caret'
##
## The following object is masked from 'package:purrr':
##
##     lift
```

```
#library(DMwR)
library(corrplot)
```

```
## corrplot 0.92 loaded
```

```
# Load teh data set
# read the hospital data
diabetic <- read.csv("https://raw.githubusercontent.com/joewarner89/CUNY-607/main/Final%20Project/diabetic_data.csv", string
sAsFactors = F)
data_info <- read.csv("https://raw.githubusercontent.com/joewarner89/CUNY-607/main/Final%20Project/IDS_mapping.csv", strings
AsFactors = F)

# Let map some column
admin <- data_info %>% select(1,2) %>% stats::na.omit()
disch <- data_info %>% select(3,4) %>% na.omit()
admin_sc <- data_info %>% select(5,6) %>% na.omit()

# Remapping the column name with id

diabetic <- diabetic %>% inner_join(admin,by =  "admission_type_id") %>%
  inner_join(disch,by = "discharge_disposition_id") %>%
  inner_join(admin_sc,by = "admission_source_id") %>%
  select(1:5,7,admission_type_name,discharge_disposition_id,discharge_disposition_name
        ,admission_source_id,admission_source_name,9:51)
# get a look on how the data set work
skim(diabetic)
```

Data summary

Name	diabetic
Number of rows	101766
Number of columns	52
_____	
Column type frequency:	
character	39
numeric	13
_____	
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
race	0	1	1	15	0	6	0
gender	0	1	4	15	0	3	0
age	0	1	6	8	0	10	0
admission_type_name	0	1	4	13	0	8	0
discharge_disposition_name	0	1	4	105	0	26	0
admission_source_name	0	1	4	58	0	17	0
payer_code	0	1	1	2	0	18	0
medical_specialty	0	1	1	36	0	73	0
diag_1	0	1	1	6	0	717	0
diag_2	0	1	1	6	0	749	0
diag_3	0	1	1	6	0	790	0
max_glu_serum	0	1	4	4	0	4	0
A1Cresult	0	1	2	4	0	4	0
metformin	0	1	2	6	0	4	0
repaglinide	0	1	2	6	0	4	0
nateglinide	0	1	2	6	0	4	0

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
chlorpropamide	0	1	2	6	0	4	0
glimepiride	0	1	2	6	0	4	0
acetohexamide	0	1	2	6	0	2	0
glipizide	0	1	2	6	0	4	0
glyburide	0	1	2	6	0	4	0
tolbutamide	0	1	2	6	0	2	0
pioglitazone	0	1	2	6	0	4	0
rosiglitazone	0	1	2	6	0	4	0
acarbose	0	1	2	6	0	4	0
miglitol	0	1	2	6	0	4	0
troglitazone	0	1	2	6	0	2	0
tolazamide	0	1	2	6	0	3	0
examide	0	1	2	2	0	1	0
citoglipton	0	1	2	2	0	1	0
insulin	0	1	2	6	0	4	0
glyburide.metformin	0	1	2	6	0	4	0
glipizide.metformin	0	1	2	6	0	2	0
glimepiride.pioglitazone	0	1	2	6	0	2	0
metformin.rosiglitazone	0	1	2	6	0	2	0
metformin.pioglitazone	0	1	2	6	0	2	0
change	0	1	2	2	0	2	0
diabetesMed	0	1	2	3	0	2	0
readmitted	0	1	2	3	0	3	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
encounter_id	0	1	165201645.62	102640295.98	12522	84961194	152388987	230270888	443867222	
patient_nbr	0	1	54330400.69	38696359.35	135	23413221	45505143	87545950	189502619	
admission_type_id	0	1	2.02	1.45	1	1	1	3	8	
discharge_disposition_id	0	1	3.72	5.28	1	1	1	4	28	
admission_source_id	0	1	5.75	4.06	1	1	7	7	25	
time_in_hospital	0	1	4.40	2.99	1	2	4	6	14	
num_lab_procedures	0	1	43.10	19.67	1	31	44	57	132	
num_procedures	0	1	1.34	1.71	0	0	1	2	6	
num_medications	0	1	16.02	8.13	1	10	15	20	81	
number_outpatient	0	1	0.37	1.27	0	0	0	0	42	
number_emergency	0	1	0.20	0.93	0	0	0	0	76	
number_inpatient	0	1	0.64	1.26	0	0	0	1	21	
number_diagnoses	0	1	7.42	1.93	1	6	8	9	16	

Understanding the data set

According the UC Irvine, The dataset represents ten years (1999-2008) of clinical care at 130 US hospitals and integrated delivery networks. It includes over 50 features representing patient and hospital outcomes. Information was extracted from the database for encounters that satisfied the following criteria. (1) It is an inpatient encounter (a hospital admission). (2) It is a diabetic encounter, that is, one during which any kind of diabetes was entered into the system as a diagnosis. (3) The length of stay was at least 1 day and at most 14 days. (4) Laboratory tests were performed during the encounter. (5) Medications were administered during the encounter.

There is a lot of categorical values that are significant to a hospital encounter and readmission in this data set. I need to remove any undocumented observations and mark them NA so later on it doesn't affect the classification model.

```
##               encounter_id                patient_nbr
##                          0                          0
##                       race                     gender
##                       2273                          0
##                        age          admission_type_id
##                          0                          0
##        admission_type_name   discharge_disposition_id
##                          0                          0
## discharge_disposition_name        admission_source_id
##                          0                          0
##      admission_source_name           time_in_hospital
##                          0                          0
##                  payer_code          medical_specialty
##                      40256                      49949
##         num_lab_procedures             num_procedures
##                          0                          0
##            num_medications          number_outpatient
##                          0                          0
##           number_emergency           number_inpatient
##                          0                          0
##                     diag_1                     diag_2
##                         21                        358
##                     diag_3           number_diagnoses
##                       1423                          0
##              max_glu_serum                  A1Cresult
##                          0                          0
##                 metformin                repaglinide
##                          0                          0
##               nateglinide             chlorpropamide
##                          0                          0
##               glimepiride              acetohexamide
##                          0                          0
##                 glipizide                 glyburide
##                          0                          0
##               tolbutamide              pioglitazone
##                          0                          0
##             rosiglitazone                  acarbose
##                          0                          0
##                  miglitol              troglitazone
##                          0                          0
##                tolazamide                    examide
##                          0                          0
##               citoglipton                    insulin
##                          0                          0
##        glyburide.metformin        glipizide.metformin
##                          0                          0
##   glimepiride.pioglitazone    metformin.rosiglitazone
##                          0                          0
##     metformin.pioglitazone                     change
##                          0                          0
##                diabetesMed                readmitted
##                          0                          0
```

```
##               encounter_id                patient_nbr
##                          0                          0
##                       race                     gender
##                          0                          0
##                        age          admission_type_id
##                          0                          0
##        admission_type_name   discharge_disposition_id
##                          0                          0
## discharge_disposition_name        admission_source_id
##                          0                          0
##      admission_source_name           time_in_hospital
##                          0                          0
##                 payer_code          medical_specialty
##                          0                          0
##         num_lab_procedures             num_procedures
##                          0                          0
##            num_medications          number_outpatient
##                          0                          0
##           number_emergency           number_inpatient
##                          0                          0
##                     diag_1                     diag_2
##                          0                          0
##                     diag_3           number_diagnoses
##                          0                          0
##              max_glu_serum                 A1Cresult
##                          0                          0
##                  metformin               repaglinide
##                          0                          0
##                nateglinide             chlorpropamide
##                          0                          0
##                glimepiride              acetohexamide
##                          0                          0
##                  glipizide                 glyburide
##                          0                          0
##                tolbutamide              pioglitazone
##                          0                          0
##              rosiglitazone                  acarbose
##                          0                          0
##                   miglitol              troglitazone
##                          0                          0
##                 tolazamide                    examide
##                          0                          0
##                citoglipton                    insulin
##                          0                          0
##        glyburide.metformin        glipizide.metformin
##                          0                          0
##   glimepiride.pioglitazone    metformin.rosiglitazone
##                          0                          0
##     metformin.pioglitazone                     change
##                          0                          0
##                 diabetesMed                readmitted
##                          0                          0
```

```
## [1] 25926    52
```

According to our summary, some data columns have only unique categorical value. Single value columns need to be removed in the data set.

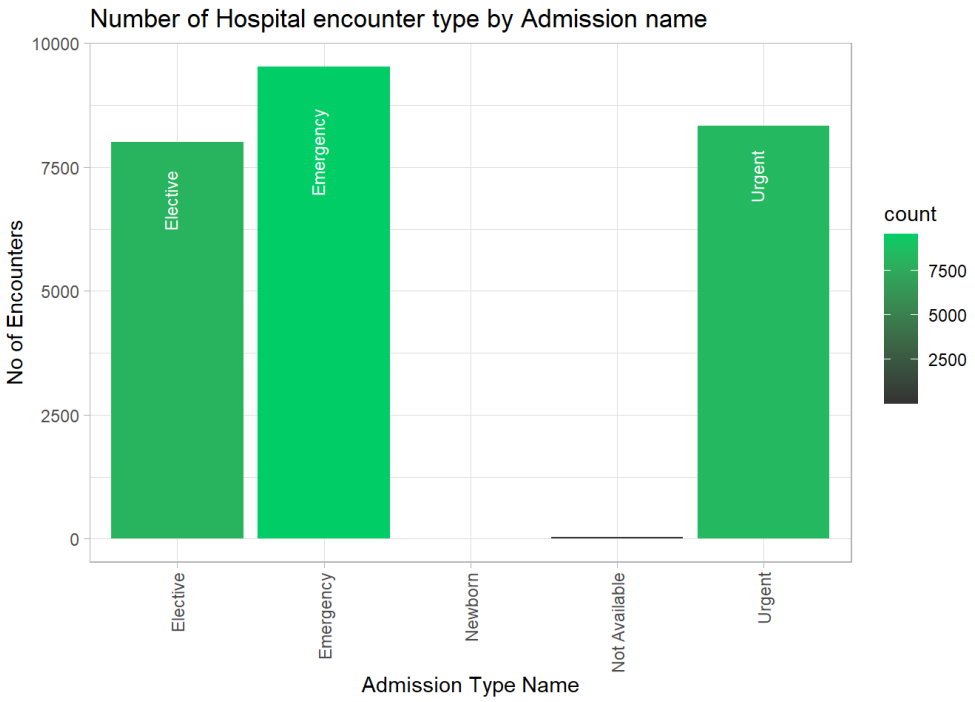
```
diabetic %>% select(citoglipton,examide,troglitazone,glimepiride.pioglitazone,
                    metformin.rosiglitazone,acetohexamide,tolbutamide) %>%
  distinct()
```

```
##   citoglipton examide troglitazone glimepiride.pioglitazone
## 1          No      No           No                       No
##   metformin.rosiglitazone acetohexamide tolbutamide
## 1                      No            No          No
```

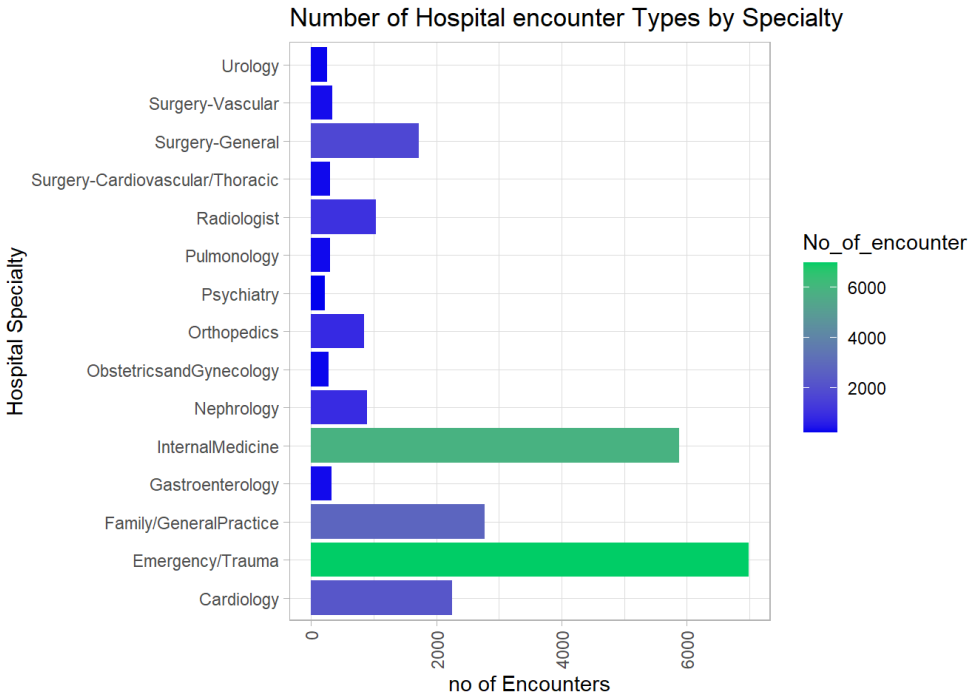
```
#remove columns
diabetic$citoglipton <- NULL
diabetic$examide <- NULL
diabetic$troglitazone <- NULL
diabetic$glimepiride.pioglitazone <- NULL
diabetic$acetohexamide <- NULL
diabetic$metformin.rosiglitazone <- NULL
diabetic$tolbutamide <- NULL

dim(diabetic)
```

```
## [1] 25926    45
```

```
#### Number of encounter type by Admission name
diabetic %>%  group_by(admission_type_name) %>% summarise(count = n()) %>% group_by(admission_type_name) %>%

  ggplot( mapping = aes(x=admission_type_name, y=count, fill=count))+
  geom_bar(position = "dodge", stat="identity")+
  scale_fill_gradient(low = "grey20", high = "springgreen3")+
  theme_light()+
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1))+labs(height=10, width=5)+
  geom_text(aes(label = admission_type_name),color="white",hjust= 1.5, vjust = 0, size = 3, angle = 90, position = position_
dodge(width = 1)) + labs(title = "Number of Hospital encounter type by Admission name", x = "Admission Type Name", y = "No o
f Encounters")
```




```
# number of encounters by
diabetic  %>% group_by(specialty = medical_specialty) %>%
  summarise(No_of_encounter = n()) %>%
  arrange(desc(No_of_encounter)) %>%
  top_n(15, wt = No_of_encounter) %>%
  ggplot(aes(x = specialty, y = No_of_encounter, fill = No_of_encounter)) +
  geom_bar(position = "dodge", stat="identity")+
  scale_fill_gradient(low = "blue2", high = "springgreen3")+
  theme_light()+
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1))+labs(height=10, width=5)+
  coord_flip() + labs(title = "Number of Hospital encounter Types by Specialty", x = "Hospital Specialty", y = "no of Encoun
ters")
```




```
# remove the icd 9 code to reveal more inormation

diabetic$diag_1 <- str_replace_all(diabetic$diag_1, "[[:punct:]]", "")
diabetic$diag_2 <- str_replace_all(diabetic$diag_2, "[[:punct:]]", "")
diabetic$diag_3 <- str_replace_all(diabetic$diag_3, "[[:punct:]]", "")

# read the icd 9 code from
icd_9 <- read.delim("https://raw.githubusercontent.com/joewarner89/CUNY-607/main/Final%20Project/icd9.txt",stringsAsFactors
= F) %>% rename(diag_1 = 1,Short_desc = 3)
```

The risk of developing pneumonia increase with diabetes. According to healthline.com, diabetes weaken your immune system. In our data set, we have 908 Hospital visits that result in a claim for pneumonia and about 185 for population ages 18 years and older that was diagnosed with uncontrolled diabetes. We may have more patients that fits these diagnosis.

The data do not contain any diagnosis name and group name. We download the ICD 9 code from CMS data and major name was collected in this link https://juniperpublishers.com/ctbeb/CTBEB.MS.ID.555715.php (https://juniperpublishers.com/ctbeb/CTBEB.MS.ID.555715.php)

The next step is to create diagnosis category to capture all major events. see picture below:

Table 2: Values for diagnosis in the final dataset. In the analysis, groups that covered less than 3.5% of encounters were grouped into the "other" category.

Group Name	ICD-9 Codes	Descriptions
Circulatory	390-459,785	Diseases of the circulatory system
Respiratory	460-519,786	Diseases of the respiratory system
Digestive	520-579,787	Diseases of the digestive system
Diabetes	250.xx	Diabetes mellitus
Injury	800-999	Injury and poisoning
Musculoskeletal	710-739	Diseases of the musculoskeletal system and connective tissue
Genitourinary	580-629,788	Diseases of the genitourinary system
Neoplasms	140-239	Neoplasms
	780,781,784,790-799	Other symptoms, signs, and ill-defined conditions
	240-279, excluding 250	Endocrine, nutritional, and metabolic diseases and immunity disorders, without diabetes
	680-709,782	Diseases of the skin and subcutaneous tissue
	001-139	Infectious and parasitic diseases
Other	290-319	Mental disorders
	E-V	External causes of injury an supplemental classification
	280-289	Diseases of the blood and blood-forming organs
	320-359	Disease of the nervous system
	630-679	Complications of pregnancy, childbirth, and the puerperium
	360=389	Diseases of the sense organs
	740-759	Congenital anomalies

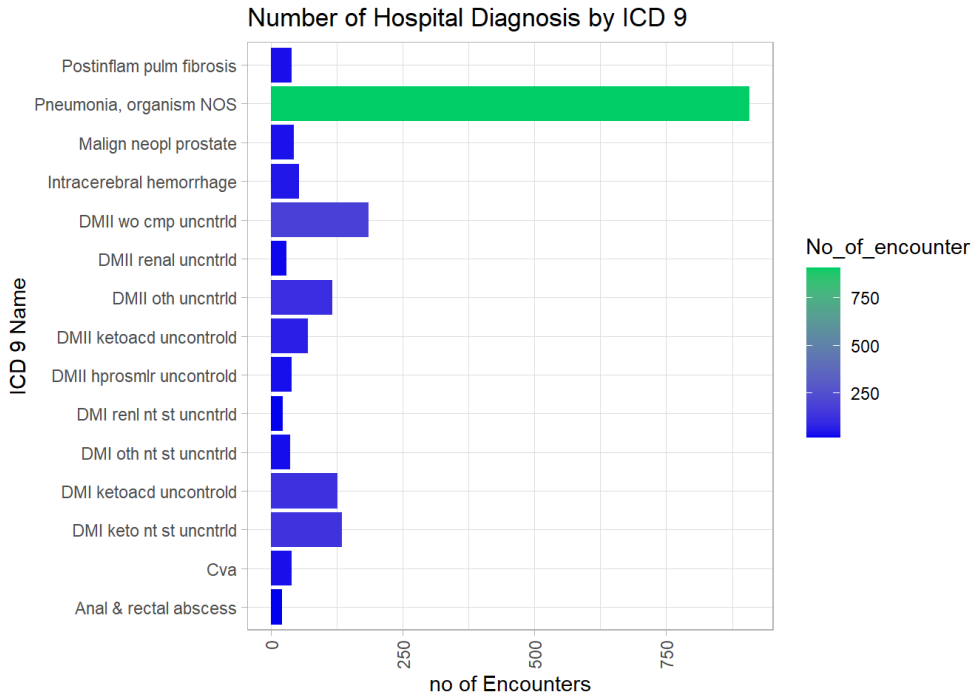
Preliminary analysis and the final dataset

Our analysis demonstrates that there are unique diabetic readmission profiles within the following age groups:

Age Group	Description
[0-30)	From 0 to 29 years old
[30-70)	From 30 to 69 years old
[70-100)	From 70 to 99 years old

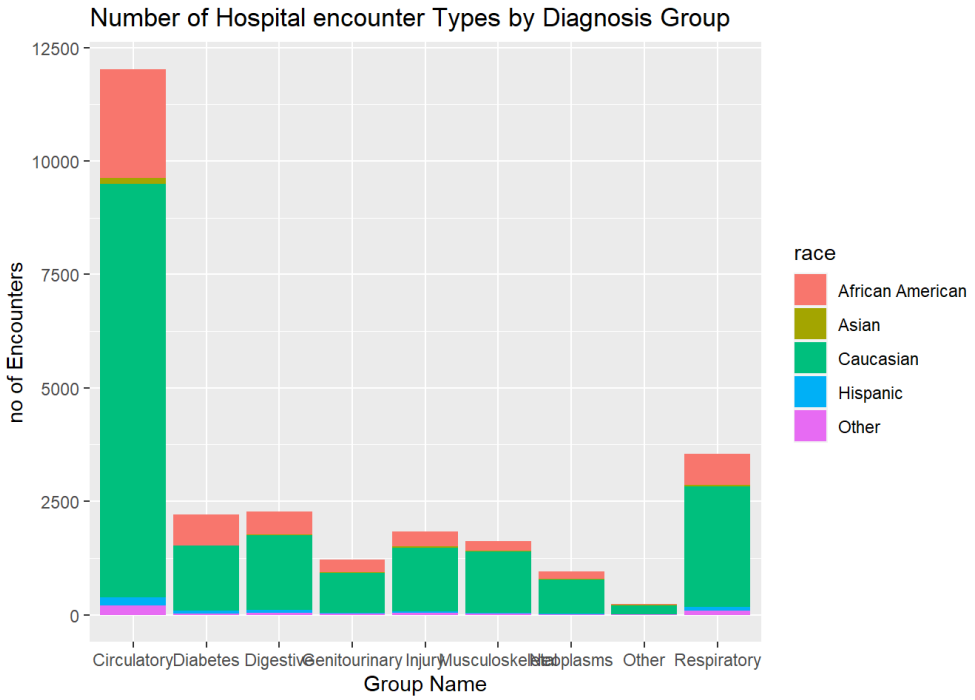
Main Group Name for ICD 9

```
# look at the main diagnosis for

diabetic   %>%
  inner_join(icd_9,by="diag_1")  %>% select(encounter_id,diag_1,Short_desc) %>%
  group_by(Short_desc) %>%  summarize(No_of_encounter = n()) %>%
  arrange(desc(No_of_encounter)) %>%
  top_n(15, wt = No_of_encounter) %>%
  ggplot(aes(x = Short_desc, y = No_of_encounter, fill = No_of_encounter)) +
  geom_bar(position = "dodge", stat="identity")+
  scale_fill_gradient(low = "blue2", high = "springgreen3")+
  theme_light()+
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1))+labs(height=10, width=5)+
  coord_flip() + labs(title = "Number of Hospital Diagnosis by ICD 9", x = "ICD 9 Name", y = "no of Encounters")
```



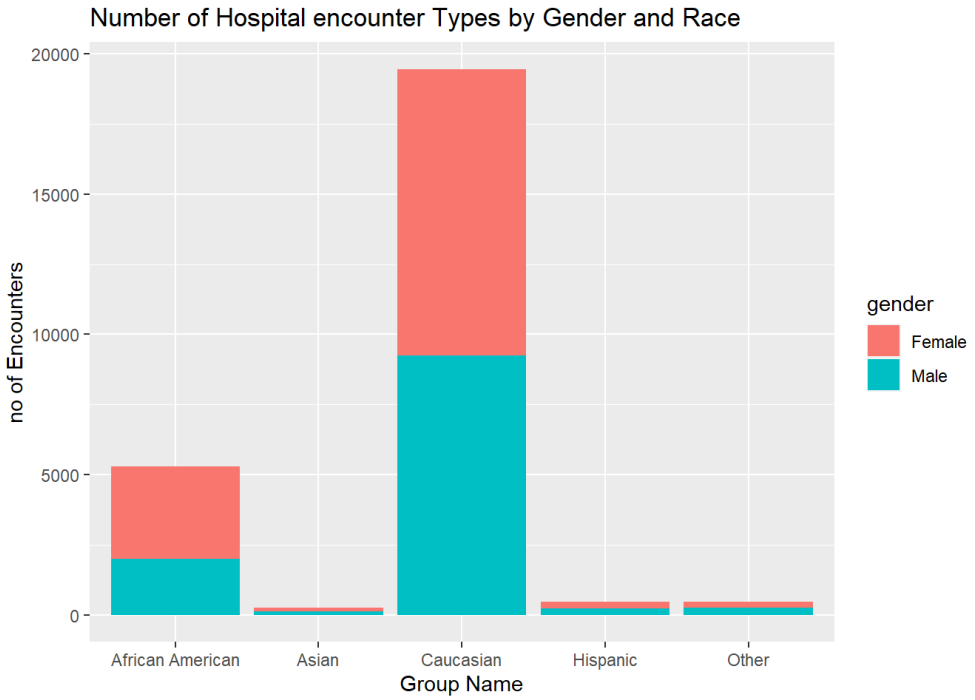

```
# Diabetes Poor controls
diag_cate <- c("2511", "2512", "24901", "24930", "24931", "24941", "24951"," 24961", '24971', "24981", "24991", "25002", "25003", "25030", "25031",
               "25032", "25033", "25042", "25043,", "25052", "25053", "25062", "25063", "25072", "25073", "25082", "25083", "25092", "25093")

# Create diagnosis table
data <- mutate(diabetic, diagnosis =
         ifelse(str_detect(diag_1, "V") | str_detect(diag_1, "E"),"Other",
                # disease codes starting with V or E are in "other" category;
                ifelse(str_detect(diag_1, "250"), "Diabetes",
                       ifelse((as.integer(diag_1) %in% diag_cate & as.integer(diag_1) <= 459) | as.integer(diag_1) == 785, "Circulatory",
                              ifelse((as.integer(diag_1) >= 460 & as.integer(diag_1) <= 519) | as.integer(diag_1) == 786, "Respiratory",
                                     ifelse((as.integer(diag_1) >= 520 & as.integer(diag_1) <= 579) | as.integer(diag_1) == 787, "Digestive",
                                            ifelse((as.integer(diag_1) >= 580 & as.integer(diag_1) <= 629) | as.integer(diag_1) == 788, "Genitourinary",
                                                   ifelse((as.integer(diag_1) >= 140 & as.integer(diag_1) <= 239), "Neoplasms",
                                                          ifelse((as.integer(diag_1) >= 710 & as.integer(diag_1) <= 739), "Musculoskeletal",
                                                                 ifelse((as.integer(diag_1) >= 800 & as.integer(diag_1) <= 999), "Injury",
                                                                        "Circulatory")))))))))


# Group Name by encounter count

data %>%  group_by(diagnosis,race) %>% summarise(count = n()) %>% group_by(diagnosis) %>%
  ggplot(aes(x =diagnosis, y = count, fill = race)) +
  geom_col() +
   labs(title = "Number of Hospital encounter Types by Diagnosis Group", x = "Group Name", y = "no of Encounters")
```

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




```
# Number of men and women affected by diabetes
data %>%  group_by(race,gender) %>% summarise(count = n()) %>%  arrange(desc(count)) %>%
  ggplot(aes(x =race, y = count, fill = gender)) +
  geom_col() +
  labs(title = "Number of Hospital encounter Types by Gender and Race", x = "Group Name", y = "no of Encounters")
```

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




```
# gender stats

gender_stat <- data %>%  group_by(race,gender) %>% summarise(count = n()) %>%  arrange(desc(count))
```

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

```
htmlTable(gender_stat)
```

	race	gender	count
1	Caucasian	Female	10221
2	Caucasian	Male	9240
3	African American	Female	3307
4	African American	Male	1980
5	Other	Male	247
6	Hispanic	Female	237
7	Hispanic	Male	227
8	Other	Female	218
9	Asian	Male	130
10	Asian	Female	119

Diabetes mellitus or diabetes disproportionately affects minority populations. Most patients who poor control over their diabetes treatment face circulatory issues. In the data set, both caucasian and African american have the high number of circulatory diagnosis. There are more women caucasian that are diagnosed with diabetes more than every ethnic group.

Multivariate and Univariate Analysis

We modify the input of some the features to scale the data set. I make change to some categorical values. This data set has more than 15 categorical values. It is very useful to understand the impact of each variables in real life. Medication play a big part of treatments for Diabetes. Some of the features present a challenge due to the way they were created. Variables with low variances are not reliable for statistical modeling.

We combine all the low variance columns to add more variances to the data and preserve the information. Based on the initial graph we did, we need to remove outliers.

```
# reassign data set
lib <- data

## Removing duplicate patients encounter
lib <- lib[!duplicated(lib$patient_nbr),]

# Change some categorical data to numerical
# recategorize the column age
lib$age <- ifelse(lib$age == "[0-10)",  0, lib$age);
lib$age <- ifelse(lib$age == "[10-20)", 10, lib$age);
lib$age <- ifelse(lib$age == "[20-30)", 20, lib$age);
lib$age <- ifelse(lib$age == "[30-40)", 30, lib$age);
lib$age <- ifelse(lib$age == "[40-50)", 40, lib$age);
lib$age <- ifelse(lib$age == "[50-60)", 50, lib$age);
lib$age <- ifelse(lib$age == "[60-70)", 60, lib$age);
lib$age <- ifelse(lib$age == "[70-80)", 70, lib$age);
lib$age <- ifelse(lib$age == "[80-90)", 80, lib$age);
lib$age <- ifelse(lib$age == "[90-100)", 90, lib$age);

# Change categorical values
lib$max_glu_serum <- ifelse(lib$max_glu_serum == "None",  0, lib$max_glu_serum);
lib$max_glu_serum <- ifelse(lib$max_glu_serum == "Norm",  100, lib$max_glu_serum);
lib$max_glu_serum <- ifelse(lib$max_glu_serum == ">200",  200, lib$max_glu_serum);
lib$max_glu_serum <- ifelse(lib$max_glu_serum == ">300",  300, lib$max_glu_serum);

lib$A1Cresult <- ifelse(lib$A1Cresult == "None",  0, lib$A1Cresult);
lib$A1Cresult <- ifelse(lib$A1Cresult == "Norm",  5, lib$A1Cresult);
lib$A1Cresult <- ifelse(lib$A1Cresult == ">7",    7, lib$A1Cresult);
lib$A1Cresult <- ifelse(lib$A1Cresult == ">8",    8, lib$A1Cresult);

lib$metformin <- ifelse(lib$metformin == 'No', 0 , lib$metformin);
lib$metformin <- ifelse(lib$metformin == 'Steady', 1 , lib$metformin);
lib$metformin <- ifelse(lib$metformin == 'Up', 2 , lib$metformin);
lib$metformin <- ifelse(lib$metformin == 'Down', 3 , lib$metformin);

lib$repaglinide <- ifelse(lib$repaglinide == 'No', 0 , lib$repaglinide);
lib$repaglinide <- ifelse(lib$repaglinide == 'Steady', 1 , lib$repaglinide);
lib$repaglinide <- ifelse(lib$repaglinide == 'Up', 2 , lib$repaglinide);
lib$repaglinide <- ifelse(lib$repaglinide == 'Down', 3 , lib$repaglinide);

lib$chlorpropamide <- ifelse(lib$chlorpropamide == 'No', 0 , lib$chlorpropamide);
lib$chlorpropamide <- ifelse(lib$chlorpropamide == 'Steady', 1 , lib$chlorpropamide);
lib$chlorpropamide <- ifelse(lib$chlorpropamide == 'Up', 2 , lib$chlorpropamide);
lib$chlorpropamide <- ifelse(lib$chlorpropamide == 'Down', 3 , lib$chlorpropamide);

lib$nateglinide <- ifelse(lib$nateglinide == 'No', 0 , lib$nateglinide);
lib$nateglinide <- ifelse(lib$nateglinide == 'Steady', 1 , lib$nateglinide);

lib$glimepiride <- ifelse(lib$glimepiride == 'No', 0 , lib$glimepiride);
lib$glimepiride <- ifelse(lib$glimepiride == 'Steady', 1 , lib$glimepiride);
lib$glimepiride <- ifelse(lib$glimepiride == 'Up', 2 , lib$glimepiride);

lib$pioglitazone <- ifelse(lib$pioglitazone == 'No', 0 , lib$pioglitazone);
lib$pioglitazone <- ifelse(lib$pioglitazone == 'Steady', 1 , lib$pioglitazone);
lib$pioglitazone <- ifelse(lib$pioglitazone == 'Up', 2 , lib$pioglitazone);

lib$glyburide <- ifelse(lib$glyburide == 'No', 0 , lib$glyburide);
lib$glyburide <- ifelse(lib$glyburide == 'Steady', 1 , lib$glyburide);
lib$glyburide <- ifelse(lib$glyburide == 'Up', 2 , lib$glyburide);
lib$glyburide <- ifelse(lib$glyburide == 'Down', 3 , lib$glyburide);

lib$acarbose <- ifelse(lib$acarbose == 'No', 0 , lib$acarbose);
lib$acarbose <- ifelse(lib$acarbose == 'Steady', 1 , lib$acarbose);

lib$insulin <- ifelse(lib$insulin == 'No', 0 , lib$insulin);
lib$insulin <- ifelse(lib$insulin == 'Steady', 1 , lib$insulin);
lib$insulin <- ifelse(lib$insulin == 'Up', 2 , lib$insulin);
lib$insulin <- ifelse(lib$insulin == 'Down', 3 , lib$insulin);

lib$glipizide.metformin <- ifelse(lib$glipizide.metformin == 'No', 0 , lib$glipizide.metformin);
```

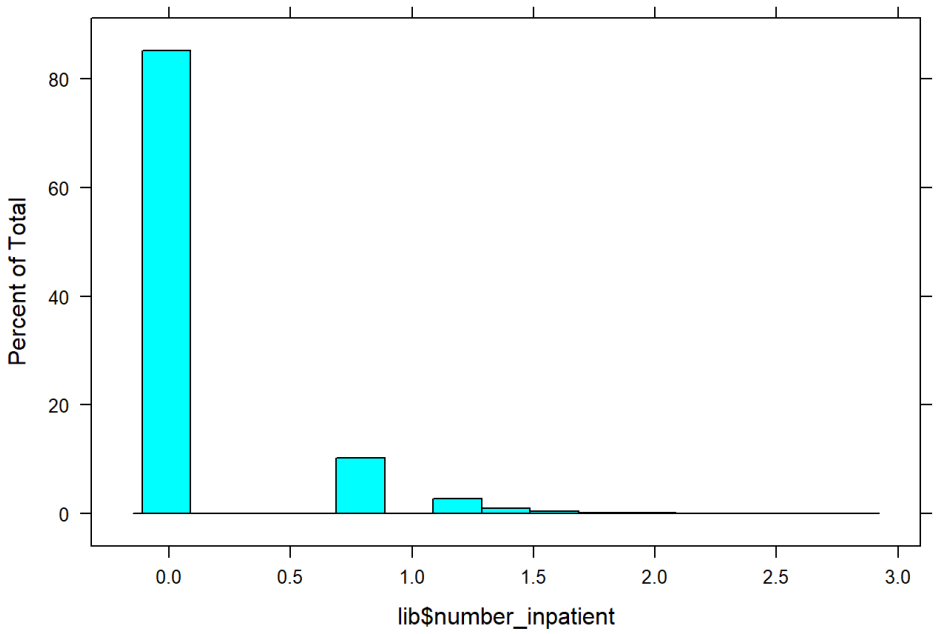
```
lib$glipizide.metformin <- ifelse(lib$glipizide.metformin == 'Steady', 1 , lib$glipizide.metformin);

lib$change <- ifelse(lib$change == 'No', 0 , lib$change);
lib$change <- ifelse(lib$change == 'Ch', 1 , lib$change);

lib$diabetesMed <- ifelse(lib$diabetesMed == 'No', 0 , lib$diabetesMed);
lib$diabetesMed <- ifelse(lib$diabetesMed == 'Yes', 1 , lib$diabetesMed);
#See Variable with low variances
nearZeroVar(lib, names = T, freqCut = 19, uniqueCut = 10)
```

```
##  [1] "max_glu_serum"          "repaglinide"            "nateglinide"
##  [4] "chlorpropamide"         "acarbose"               "miglitol"
##  [7] "tolazamide"             "glyburide.metformin"    "glipizide.metformin"
## [10] "metformin.pioglitazone"
```

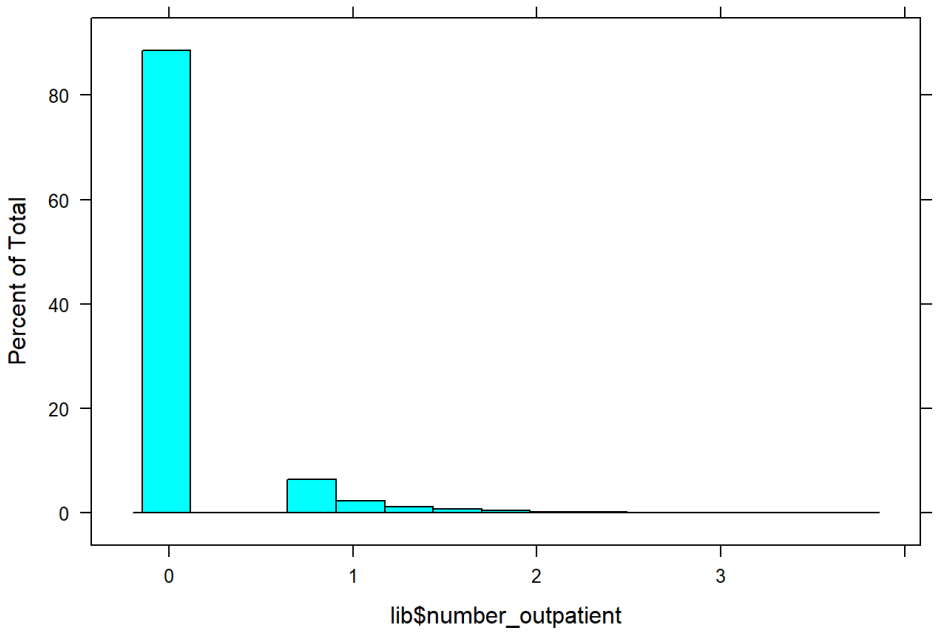
```
# re categorize encounter
# Encounter is unique for every visit, so we are going create visit column to capture
# the number of outpatient inpatient and emergency

lib$visits = lib$number_outpatient + lib$number_emergency + lib$number_inp
readmitted = lib$readmitted
lib <- subset(lib, select =-c(readmitted))
lib$readmitted = readmitted

# identify low variance in the data set
#This column has low variances
keys <- nearZeroVar(lib, names = T, freqCut = 19, uniqueCut = 10)
keys
```

```
##  [1] "max_glu_serum"          "repaglinide"            "nateglinide"
##  [4] "chlorpropamide"         "acarbose"               "miglitol"
##  [7] "tolazamide"             "glyburide.metformin"    "glipizide.metformin"
## [10] "metformin.pioglitazone"
```

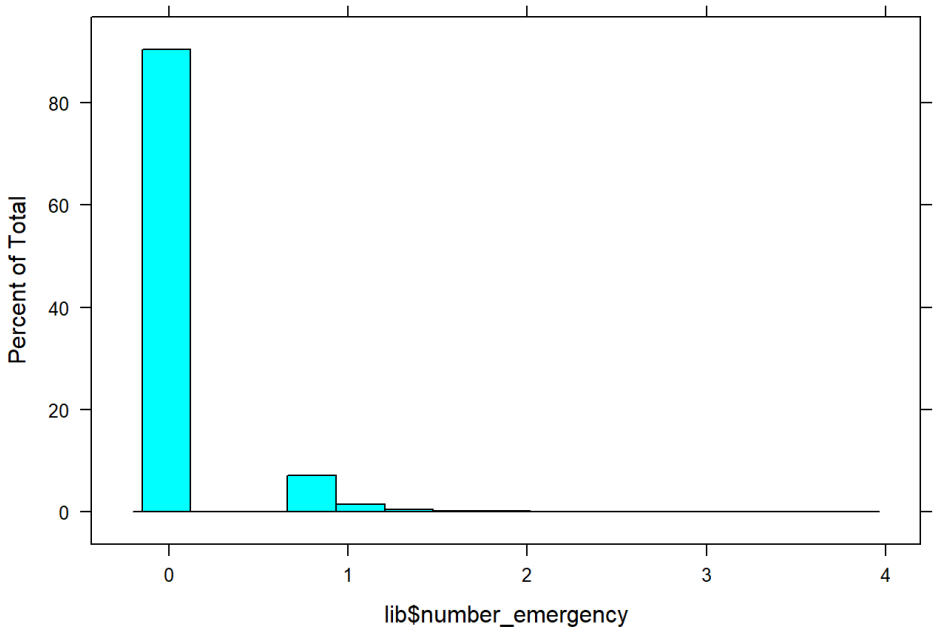
```
# Low variance is usefull to identify outliers
# correlation would be lower if variance is low.
lib$num_med <- 0
lib$num_changes <- 0
for(key in keys){
  lib$num_med <- ifelse(lib[key] != 0, lib$num_med + 1, lib$num_med)
  lib$num_changes <- ifelse((lib[key] == 1 | lib[key] == 2 | lib[key] == 3), lib$num_changes + 1, lib$num_changes)
}



## Normalize, Remove Outliers, and Standardize Numerical Features
lib$number_inpatient <- log1p(lib$number_inpatient)
lib$number_outpatient <- log1p(lib$number_outpatient)
lib$number_emergency <- log1p(lib$number_emergency)

histogram(lib$number_inpatient)
```




```
histogram(lib$number_outpatient)
```




```
histogram(lib$number_emergency)
```



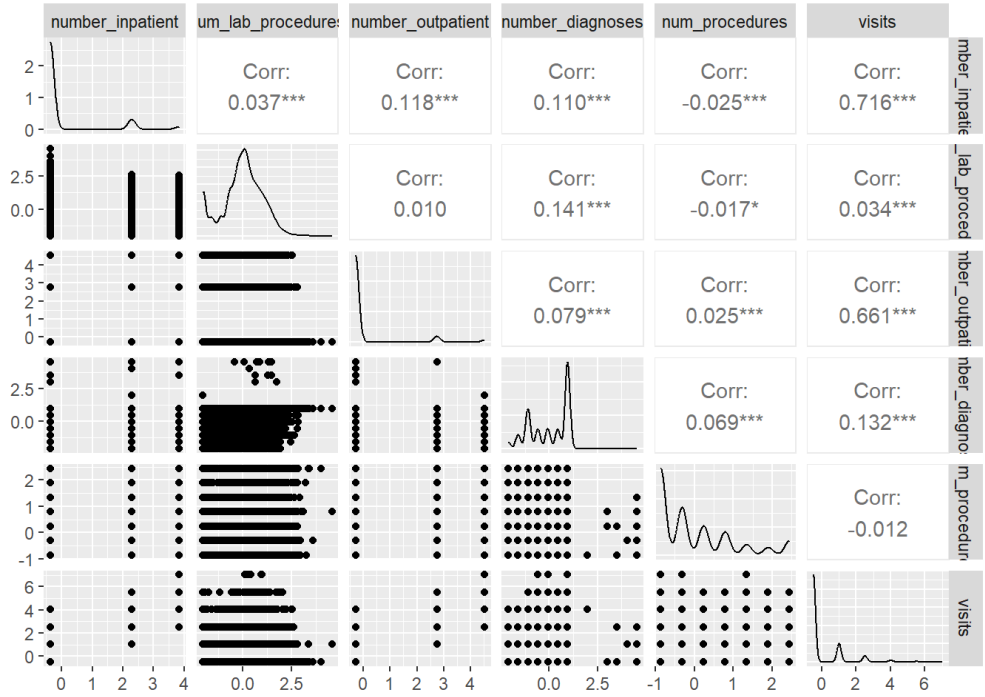

```
non_outliers = function(x, zs) {
  temp <- (x - mean(x))/sd(x)
  return(temp < zs)
}



lib <- lib[non_outliers(lib$number_inpatient, 3),]
lib <- lib[non_outliers(lib$number_outpatient, 3),]
lib <- lib[non_outliers(lib$number_emergency, 3),]
lib <- subset(lib, select = -c(number_emergency))



#Normalise skewed features and removing outliers using z-score

cols <- dplyr::select_if(lib, is.numeric)
temp <- scale(dplyr::select_if(lib, is.numeric))
for(col in colnames(cols)){
  lib[,col] <- temp[,col]
}
str(lib)
```

```
## 'data.frame':    17547 obs. of  48 variables:
##  $ encounter_id              : num  -1.34 -1.33 -1.33 -1.32 -1.23 ...
##  $ patient_nbr               : num  -0.945 -0.938 -0.938 -0.934 -1.286 ...
##  $ race                      : chr  "Caucasian" "Caucasian" "Caucasian" "Caucasian" ...
##  $ gender                    : chr  "Female" "Female" "Female" "Male" ...
##  $ age                       : chr  "70" "60" "90" "70" ...
##  $ admission_type_id         : num  -1.1297 0.0569 -1.1297 -1.1297 -1.1297 ...
##  $ admission_type_name       : chr  "Emergency" "Urgent" "Emergency" "Emergency" ...
##  $ discharge_disposition_id  : num  4.4708 -0.4262 -0.4262 -0.193 0.0402 ...
##  $ discharge_disposition_name: chr  "Discharged/transferred to another rehab fac including rehab units of a hospital ."
"Discharged to home" "Discharged to home" "Discharged/transferred to another short term hospital" ...
##  $ admission_source_id       : num  0.91 -1.184 0.91 0.91 0.561 ...
##  $ admission_source_name     : chr  " Emergency Room" " Physician Referral" " Emergency Room" " Emergency Room" ...
##  $ time_in_hospital          : num  0.9463 -0.4169 -0.0761 1.9687 2.6503 ...
##  $ payer_code                : chr  "MC" "MC" "MC" "MC" ...
##  $ medical_specialty         : chr  "Orthopedics-Reconstructive" "Nephrology" "Emergency/Trauma" "InternalMedicine" ...
##  $ num_lab_procedures        : num  0.849 0.899 0.748 1.354 1.808 ...
##  $ num_procedures            : num  0.238 0.79 -0.313 -0.313 1.894 ...
##  $ num_medications           : num  -0.11 -0.557 -0.781 0.226 0.338 ...
##  $ number_outpatient         : num  -0.296 -0.296 -0.296 -0.296 -0.296 ...
##  $ number_inpatient          : num  -0.365 -0.365 -0.365 -0.365 -0.365 ...
##  $ diag_1                    : chr  "821" "V56" "532" "682" ...
##  $ diag_2                    : chr  "276" "403" "428" "427" ...
##  $ diag_3                    : chr  "285" "599" "535" "276" ...
##  $ number_diagnoses          : num  0.96 -0.58 -0.58 -0.58 -1.09 ...
##  $ max_glu_serum             : chr  "0" "0" "0" "0" ...
##  $ A1Cresult                 : chr  "0" "0" "0" "0" ...
##  $ metformin                 : chr  "0" "0" "0" "1" ...
##  $ repaglinide               : chr  "0" "0" "0" "0" ...
##  $ nateglinide               : chr  "0" "0" "0" "0" ...
##  $ chlorpropamide            : chr  "0" "0" "0" "0" ...
##  $ glimepiride               : chr  "0" "0" "1" "0" ...
##  $ glipizide                 : chr  "No" "No" "No" "No" ...
##  $ glyburide                 : chr  "0" "0" "0" "0" ...
##  $ pioglitazone              : chr  "2" "0" "0" "0" ...
##  $ rosiglitazone             : chr  "No" "No" "No" "No" ...
##  $ acarbose                  : chr  "0" "0" "0" "0" ...
##  $ miglitol                  : chr  "No" "No" "No" "No" ...
##  $ tolazamide                : chr  "No" "No" "No" "No" ...
##  $ insulin                   : chr  "1" "1" "0" "1" ...
##  $ glyburide.metformin       : chr  "No" "No" "No" "No" ...
##  $ glipizide.metformin       : chr  "0" "0" "0" "0" ...
##  $ metformin.pioglitazone    : chr  "No" "No" "No" "No" ...
##  $ change                    : chr  "1" "0" "0" "1" ...
##  $ diabetesMed               : chr  "1" "1" "1" "1" ...
##  $ diagnosis                 : chr  "Injury" "Other" "Digestive" "Circulatory" ...
##  $ visits                    : num  -0.484 -0.484 -0.484 -0.484 -0.484 ...
##  $ readmitted                : chr  "NO" "NO" "NO" "NO" ...
##  $ num_med                   : num  -0.19 -0.19 -0.19 -0.19 -0.19 ...
##  $ num_changes               : num  -0.189 -0.189 -0.189 -0.189 -0.189 ...
```

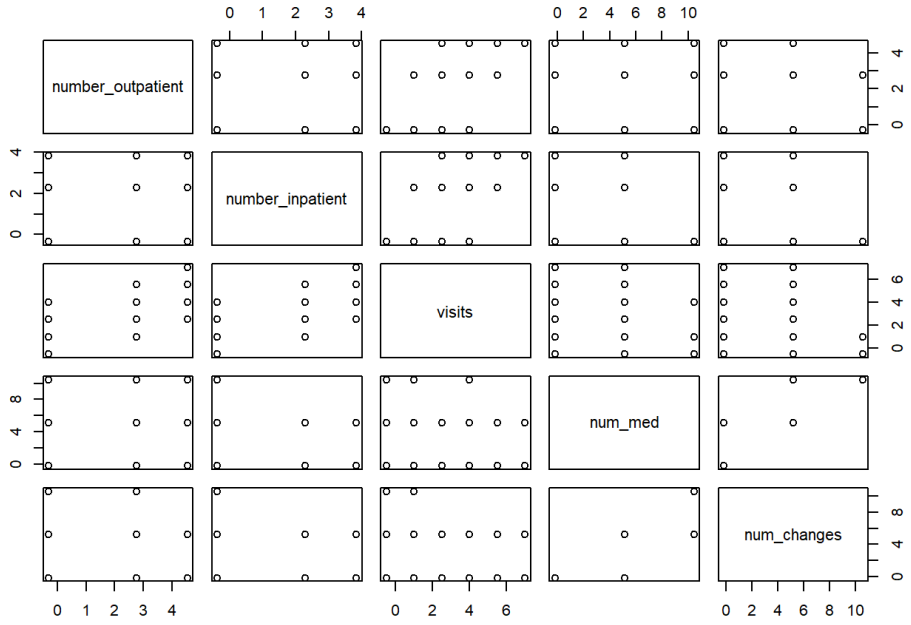
```
# see c
 # see c
library(dplyr)
# see c
library(GGally)
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```
ggpairs(lib, columns = c("number_inpatient","num_lab_procedures","number_outpatient","number_diagnoses",
                         "num_procedures","visits"))
```




```
# Change datatype
lib$num_med   <- as.numeric(lib$num_med)
# Change data type
lib$num_changes <- as.numeric(lib$num_changes)

# PLot all the variables

cor <-  lib %>% dplyr::select(number_outpatient,number_inpatient,visits, num_med, num_changes)
plot(cor)
```

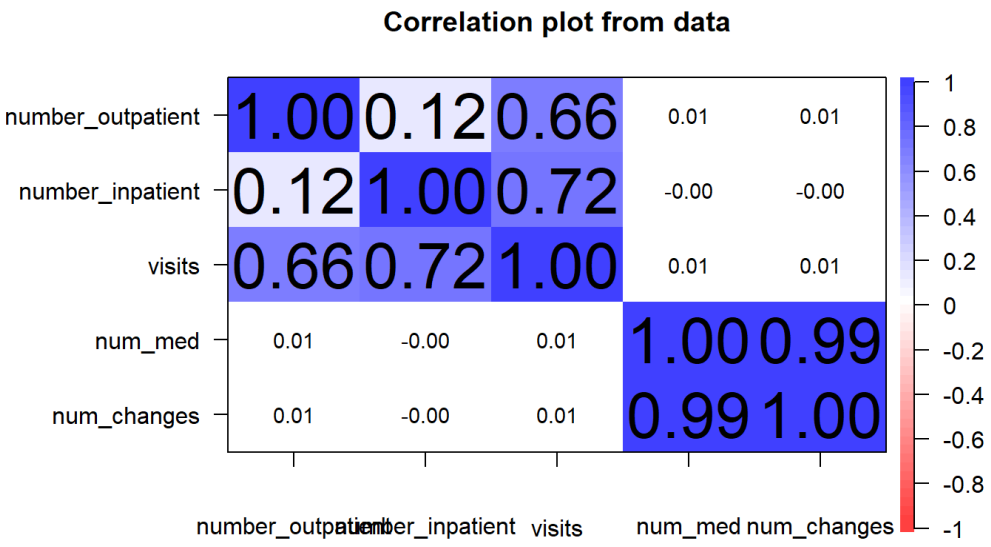



```
library(psych)
```

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

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

```
corPlot(cor)
```




```
# Model development

# Model development

# set seed
set.seed(123)
# turn row into 2 category 1 for readmitted and 0 not readmitted in the prediod of 30 days
lib$readmitted <- case_when(lib$readmitted %in% c(">30","NO") ~ "0",
                            TRUE ~ "1")

# creating  training and test
train_indices <- sample(seq_len(nrow(lib)), 0.7* nrow(lib))
train_data <- lib[train_indices, ]
test_data <- lib[-train_indices, ]

# Fit the logistic regression
model <- glm(readmitted == '1' ~  visits + number_inpatient + number_outpatient  , data = train_data)
summary(model)
```

```
##
## Call:
## glm(formula = readmitted == "1" ~ visits + number_inpatient +
##     number_outpatient, data = train_data)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -0.15409  -0.07546  -0.07546  -0.07546   0.94035
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        0.080788   0.002454  32.924   <2e-16 ***
## visits             0.005823   0.006218   0.936   0.3490
## number_inpatient   0.012472   0.004732   2.636   0.0084 **
## number_outpatient -0.006891   0.004397  -1.567   0.1171
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.07393891)
##
##     Null deviance: 911.04  on 12281  degrees of freedom
## Residual deviance: 907.82  on 12278  degrees of freedom
## AIC: 2872.1
##
## Number of Fisher Scoring iterations: 2
```

```
predicted_probs <- predict(model, newdata = test_data , type = "response")

# Convert probabilities to binary predictions
predicted_classes <- ifelse(predicted_probs > 0.1, "readmitted", "Not readmitted")
# Convert probabilities to predicted classes
##predicted_classes <- ifelse(predicted_probs > 0.5, "<30", ifelse(predicted_probs > 0.25, "NO", ">30"))


# Display the confusion matrix
confusion_matrix <- table(predicted_classes, test_data$readmitted)
print(confusion_matrix)
```

```
##
## predicted_classes    0    1
##    Not readmitted 4261  327
##    readmitted      594   83
```

```
# Calculate accuracy
accuracy <- sum(diag(confusion_matrix)) / sum(confusion_matrix)
print(paste("Accuracy:", accuracy))
```

```
## [1] "Accuracy: 0.825071225071225"
```

```
#Display the predicted labels

pred_model_1 <- test_data
pred_model_1$predictor <- predicted_classes
head(pred_model_1)
```

```
##       encounter_id patient_nbr             race gender age admission_type_id
## 24070    -1.219663   -1.100618        Caucasian Female  30       -1.12965500
## 24138    -1.217350   -1.101954        Caucasian Female  70       -1.12965500
## 24277    -1.212543   -1.180311        Caucasian Female  80       -1.12965500
## 24302    -1.211674   -1.144384        Caucasian   Male  30        0.05693921
## 24315    -1.211539   -1.066458 African American Female  80       -1.12965500
## 24345    -1.210605   -1.447905        Caucasian Female  70        0.05693921
##       admission_type_name discharge_disposition_id discharge_disposition_name
## 24070           Emergency               -0.4262194         Discharged to home
## 24138           Emergency               -0.4262194         Discharged to home
## 24277           Emergency               -0.4262194         Discharged to home
## 24302              Urgent                0.9729211                   Left AMA
## 24315           Emergency               -0.4262194         Discharged to home
## 24345              Urgent               -0.4262194         Discharged to home
##       admission_source_id                       admission_source_name
## 24070           0.5614243  Transfer from another health care facility
## 24138           0.9104398                              Emergency Room
## 24277           0.5614243  Transfer from another health care facility
## 24302          -1.1836534                          Physician Referral
## 24315           0.5614243  Transfer from another health care facility
## 24345          -1.1836534                          Physician Referral
##       time_in_hospital payer_code medical_specialty num_lab_procedures
## 24070       2.99114428         MC        Nephrology          1.7071424
## 24138      -0.75775474         MC  InternalMedicine          0.4955452
## 24277      -0.07613673         MC  InternalMedicine          0.1421626
## 24302      -0.75775474         UN  InternalMedicine          0.3440955
## 24315       0.26467227         UN  InternalMedicine          1.2023102
## 24345      -0.75775474         MC  InternalMedicine         -0.9179850
##       num_procedures num_medications number_outpatient number_inpatient diag_1
## 24070      0.2383056     0.002104749         -0.295955       -0.3652328    112
## 24138     -0.8652964    -1.005132450         -0.295955       -0.3652328    577
## 24277     -0.8652964    -0.557471472         -0.295955       -0.3652328    414
## 24302     -0.8652964    -1.117047694         -0.295955       -0.3652328    486
## 24315     -0.8652964    -1.005132450         -0.295955       -0.3652328    435
## 24345     -0.8652964    -0.781301961         -0.295955       -0.3652328    428
##       diag_2 diag_3 number_diagnoses max_glu_serum A1Cresult metformin
## 24070    996  25013        -1.093696             0         8         0
## 24138    428    414        -1.093696             0         0         0
## 24277    427    424        -1.093696             0         0         0
## 24302    493  25013        -1.093696             0         0         0
## 24315    427  25092        -1.093696             0         0         1
## 24345    250    414        -1.093696             0         0         0
##       repaglinide nateglinide chlorpropamide glimepiride glipizide glyburide
## 24070           0           0              0           0        No         0
## 24138           0           0              0           0        No         0
## 24277           0           0              0           1        No         0
## 24302           0           0              0           0        No         0
## 24315           0           0              0           0    Steady         0
## 24345           0           0              0           0        No         0
##       pioglitazone rosiglitazone acarbose miglitol tolazamide insulin
## 24070            0            No        0       No         No       1
## 24138            0            No        0       No         No       0
## 24277            0            No        0       No         No       1
## 24302            0            No        0       No         No       0
## 24315            0            No        0       No         No       0
## 24345            0            No        0       No         No       1
##       glyburide.metformin glipizide.metformin metformin.pioglitazone change
## 24070                  No                   0                      No      0
## 24138                  No                   0                      No      0
## 24277                  No                   0                      No      1
## 24302                  No                   0                      No      0
## 24315                  No                   0                      No      1
## 24345                  No                   0                      No      0
##       diabetesMed   diagnosis     visits readmitted    num_med num_changes
## 24070           1 Circulatory -0.4838539          0 -0.1898372  -0.1885303
## 24138           0   Digestive -0.4838539          0 -0.1898372  -0.1885303
## 24277           1 Circulatory -0.4838539          0 -0.1898372  -0.1885303
## 24302           0 Respiratory -0.4838539          0 -0.1898372  -0.1885303
## 24315           1 Circulatory -0.4838539          0 -0.1898372  -0.1885303
## 24345           1 Circulatory -0.4838539          0 -0.1898372  -0.1885303
```

```
##            predictor
## 24070 Not readmitted
## 24138 Not readmitted
## 24277 Not readmitted
## 24302 Not readmitted
## 24315 Not readmitted
## 24345 Not readmitted
```

```
#evaluation
##3 second model
library(boot)
```

```
##
## Attaching package: 'boot'
```

```
## The following object is masked from 'package:psych':
##
##     logit
```

```
## The following object is masked from 'package:lattice':
##
##     melanoma
```

```
# Define the logistic regression model
set.seed(123)
logistic_model <- glm(readmitted == "1" ~ visits + number_outpatient  , data = lib)
summary(logistic_model)
```

```
##
## Call:
## glm(formula = readmitted == "1" ~ visits + number_outpatient,
##     data = lib)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -0.16832  -0.07363  -0.07363  -0.07363   0.93571
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        0.079843   0.002043  39.083  < 2e-16 ***
## visits             0.021064   0.002723   7.735 1.09e-14 ***
## number_outpatient -0.013437   0.002723  -4.934 8.14e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.07323046)
##
##     Null deviance: 1289.1  on 17546  degrees of freedom
## Residual deviance: 1284.8  on 17544  degrees of freedom
## AIC: 3930.8
##
## Number of Fisher Scoring iterations: 2
```

```
# Perform 10-fold cross-validation
cv_results <- cv.glm(lib, logistic_model, K = 10)

##print(cv_results)

# Access accuracy
accuracy2 <- 1 - cv_results$delta[1]
cat("Accuracy:", accuracy2, "\n")
```

```
## Accuracy: 0.9267509
```

```
# Predict the probabilities of the positive class ("Wins")
predicted_probs2 <- predict(logistic_model, newdata = lib, type = "response")


# Convert probabilities to binary predictions
predicted_labels2 <- predicted_classes <- ifelse(predicted_probs2 > 0.1, "readmitted", "Not readmitted")

# Display results
#print(data.frame(Probabilities = predicted_probs2, Predicted_Labels = predicted_labels2))
at2 <- data.frame(Probabilities = predicted_probs2, Predicted_Labels = predicted_labels2)
head(at2)
```

```
##       Probabilities Predicted_Labels
## 20447    0.07362731   Not readmitted
## 20738    0.07362731   Not readmitted
## 20825    0.07362731   Not readmitted
## 21084    0.07362731   Not readmitted
## 23880    0.07362731   Not readmitted
## 23923    0.07362731   Not readmitted
```

```
# Display the confusion matrix
confusion_matrix2 <- table(predicted_labels2, lib$readmitted)
print(confusion_matrix2)
```

```
##
## predicted_labels2     0     1
##    Not readmitted 13617  1087
##    readmitted      2529   314
```

```
# Display the predicted labels
#print(predicted_labels2)
pred_model2 <- lib
pred_model2$predictor <- predicted_labels2
head(pred_model2)
```

```
##       encounter_id patient_nbr      race gender age admission_type_id
## 20447    -1.338575  -0.9452118 Caucasian Female  70       -1.12965500
## 20738    -1.329333  -0.9384905 Caucasian Female  60        0.05693921
## 20825    -1.326727  -0.9376905 Caucasian Female  90       -1.12965500
## 21084    -1.318555  -0.9341307 Caucasian   Male  70       -1.12965500
## 23880    -1.225515  -1.2857602 Caucasian Female  70       -1.12965500
## 23923    -1.224257  -1.0683153 Caucasian   Male  70       -1.12965500
##       admission_type_name discharge_disposition_id
## 20447           Emergency               4.47077231
## 20738              Urgent              -0.42621943
## 20825           Emergency              -0.42621943
## 21084           Emergency              -0.19302935
## 23880           Emergency               0.04016074
## 23923           Emergency               0.04016074
##                                                    discharge_disposition_name
## 20447 Discharged/transferred to another rehab fac including rehab units of a hospital .
## 20738                                                          Discharged to home
## 20825                                                          Discharged to home
## 21084                            Discharged/transferred to another short term hospital
## 23880                                                Discharged/transferred to SNF
## 23923                                                Discharged/transferred to SNF
##       admission_source_id                     admission_source_name
## 20447           0.9104398                             Emergency Room
## 20738          -1.1836534                          Physician Referral
## 20825           0.9104398                             Emergency Room
## 21084           0.9104398                             Emergency Room
## 23880           0.5614243  Transfer from another health care facility
## 23923           0.5614243  Transfer from another health care facility
##       time_in_hospital payer_code          medical_specialty num_lab_procedures
## 20447       0.94629027         MC Orthopedics-Reconstructive          0.8489277
## 20738      -0.41694574         MC                 Nephrology          0.8994109
## 20825      -0.07613673         MC           Emergency/Trauma          0.7479613
## 21084       1.96871727         MC           InternalMedicine          1.3537599
## 23880       2.65033527         UN           InternalMedicine          1.8081089
## 23923       2.65033527         MC           InternalMedicine          0.9498941
##       num_procedures num_medications number_outpatient number_inpatient diag_1
## 20447      0.2383056      -0.1098105         -0.295955       -0.3652328    821
## 20738      0.7901065      -0.5574715         -0.295955       -0.3652328    V56
## 20825     -0.3134954      -0.7813020         -0.295955       -0.3652328    532
## 21084     -0.3134954       0.2259352         -0.295955       -0.3652328    682
## 23880      1.8937085       0.3378505         -0.295955       -0.3652328    238
## 23923      1.8937085      -0.6693867         -0.295955       -0.3652328    532
##       diag_2 diag_3 number_diagnoses max_glu_serum A1Cresult metformin
## 20447    276    285        0.9603645             0         0         0
## 20738    403    599       -0.5801812             0         0         0
## 20825    428    535       -0.5801812             0         0         0
## 21084    427    276       -0.5801812             0         0         1
## 23880  25002    733       -1.0936964             0         8         0
## 23923    280    569       -1.0936964             0         0         0
##       repaglinide nateglinide chlorpropamide glimepiride glipizide glyburide
## 20447           0           0              0           0        No         0
## 20738           0           0              0           0        No         0
## 20825           0           0              0           1        No         0
## 21084           0           0              0           0        No         0
## 23880           0           0              0           0        No         0
## 23923           0           0              0           0        No         0
##       pioglitazone rosiglitazone acarbose miglitol tolazamide insulin
## 20447            2            No        0       No         No       1
## 20738            0            No        0       No         No       1
## 20825            0            No        0       No         No       0
## 21084            0            No        0       No         No       1
## 23880            0            No        0       No         No       1
## 23923            0            No        0       No         No       0
##       glyburide.metformin glipizide.metformin metformin.pioglitazone change
## 20447                  No                   0                     No      1
## 20738                  No                   0                     No      0
## 20825                  No                   0                     No      0
## 21084                  No                   0                     No      1
## 23880                  No                   0                     No      0
## 23923                  No                   0                     No      0
```

```
##       diabetesMed   diagnosis     visits readmitted    num_med num_changes
## 20447           1      Injury -0.4838539          0 -0.1898372  -0.1885303
## 20738           1       Other -0.4838539          0 -0.1898372  -0.1885303
## 20825           1   Digestive -0.4838539          0 -0.1898372  -0.1885303
## 21084           1 Circulatory -0.4838539          0 -0.1898372  -0.1885303
## 23880           1   Neoplasms -0.4838539          0 -0.1898372  -0.1885303
## 23923           0   Digestive -0.4838539          0 -0.1898372  -0.1885303
##            predictor
## 20447 Not readmitted
## 20738 Not readmitted
## 20825 Not readmitted
## 21084 Not readmitted
## 23880 Not readmitted
## 23923 Not readmitted
```

Model Evaluations

We create a multiple regression model with an accuracy 0.82 with the encoded data set. we split the data into train and test set. We notice that there are steps we can do improve the model accuracy. We add more predictors to the second model and implement a cross validation of 10 folds. We were able to achieve an accuracy of 0.92. we try to add more categorical features to the multiple regression and use cross validation of 10 folds to improve the model accuracy to 0.802. We use logistic regression to predict the categorical value of readmitted against the most correlated predictors. we achieved an accuracy of 0.627.

the second model will be selected to predict the readmission data. There are still a lot of work to get done specially in data reprocessing. We manage to achieve a high accuracy but QQ plots for every model show data are not well distributed.

Limitation

There are still outliers in the data set even we used standard deviation to be removed. Data collected should contains more relevant columns that related that are related to diabetes.

```
#####################################################
# create a new model logistic model with other predictors
set.seed(123)
logistic_model2 <-glm(readmitted== "1"  ~ age + visits + number_outpatient + max_glu_serum + insulin + diabetesMed , data =
lib)
summary(logistic_model2)
```

```
##
## Call:
## glm(formula = readmitted == "1" ~ age + visits + number_outpatient +
##     max_glu_serum + insulin + diabetesMed, data = lib)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -0.19364  -0.09269  -0.07753  -0.05726   0.99242
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -0.008688   0.155913  -0.056  0.95556
## age10             -0.006254   0.160276  -0.039  0.96888
## age20              0.039639   0.156814   0.253  0.80045
## age30              0.043047   0.156301   0.275  0.78301
## age40              0.042740   0.156063   0.274  0.78419
## age50              0.047020   0.155999   0.301  0.76311
## age60              0.063696   0.155982   0.408  0.68302
## age70              0.072645   0.155974   0.466  0.64140
## age80              0.078862   0.156000   0.506  0.61320
## age90              0.084925   0.156375   0.543  0.58708
## visits             0.020053   0.002722   7.367 1.82e-13 ***
## number_outpatient -0.012827   0.002724  -4.710 2.50e-06 ***
## max_glu_serum100  -0.082886   0.191010  -0.434  0.66434
## insulin1           0.003591   0.005739   0.626  0.53147
## insulin2           0.014393   0.007116   2.023  0.04312 *
## insulin3           0.018947   0.007150   2.650  0.00806 **
## diabetesMed1       0.024835   0.006368   3.900 9.67e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.07290601)
##
##     Null deviance: 1289.1  on 17546  degrees of freedom
## Residual deviance: 1278.0  on 17530  degrees of freedom
## AIC: 3866.9
##
## Number of Fisher Scoring iterations: 2
```

```
# Perform 10-fold cross-validation
cv_results2 <- cv.glm(lib, logistic_model2, K = 10)

head(cv_results2)
```

```
## $call
## cv.glm(data = lib, glmfit = logistic_model2, K = 10)
##
## $K
## [1] 10
##
## $delta
## [1] 0.07296679 0.07295988
##
## $seed
##   [1]       10403         624  -983674937   643431772  1162448557  -959247990
##   [7]  -133913213  2107846888   370274761 -2022780170  -412390145   848182068
##  [13]  -266662747 -1309507294  1356997179  1661823040  1749531457  -516669426
##  [19]  1042678071 -1279933428  -410084963  1151007674  -895613453  1288379032
##  [25]  -376044615 -1358274522   307686511   101447652  1796216213 -1567696558
##  [31]  1186934955 -1925339152  -472470735    80319294 -1524429145   326645436
##  [37]  -389586803  -400786966  -890731933  -852332472  1365217705 -1785317034
##  [43] -1551153185  1359863956  2098748037 -1013039742  -329721061 -1587358816
##  [49]   344102689 -1520389522   166492183  1821136236  1646453629  1056605210
##  [55] -1419044141  -806080008   520985497   711286406  2004844367 -1445006012
##  [61]  1329781621 -1188844110 -1089068661  1173875536 -1983217903   514629022
##  [67]  -237421177  -258138084  -930078099   261626442  1349308227 -1125425240
##  [73] -1677778551    25874358   409637567 -1987430924  1583257701  -136173086
##  [79]   639501307   272101120 -1024630015 -1994369842  -939499785 -1944742196
##  [85]  -591520419 -1994900358  1072996275  1119025496  2035491705 -2082894618
##  [91]   776176175   -69557596  1794806101  -178474478  -497581461   874372784
##  [97]   518669041  -370223106  1295572071 -1776240260 -1692674995  1935534762
## [103]   298421283   111542024 -1075273367   518297110  -289321569  1331379028
## [109]  1768277573  1473660482  2120850651   879016544  -864018719  1661675310
## [115]   135902679 -2136373204   735594301  1594631386  -546138989  1423929528
## [121] -1067541671  1962863430 -1923418865  -984154108  1907308341   642901618
## [127] -1095019701 -1836613104 -1171392815  1663582814 -1258689721 -2007301412
## [133]  -756910547  -712643830 -1271482109  -801485208    51646793 -1925477258
## [139] -1421379457  1104736436 -1348082651  -124611934   292791739  2126591424
## [145] -2043491647  -709285490 -1242530633  1688217996  -538353379 -1997652678
## [151]   -48432781   575772696   942146361    57506214  -948054033   -72610460
## [157]  1389939989   656100050   -25586645 -2012424848  1854773937  1391516862
## [163] -2100008409  -140248004 -1638135795 -2077746326  -118729245 -1417654840
## [169]   662270249   942125782 -1363864737   744183316  2123821573   -80802046
## [175] -1753997669  1277518112  1090348705  1338137582   423408535   -28214548
## [181]  1164536573  1524008346   673959507   853634936 -1599644903 -2135083002
## [187]  -345756977 -1070478652   971985653  -556736718  -406174453   663083216
## [193]  1258368657  1306568478  1820350727 -1068259940  -402617875  1499233226
## [199] -1121819965 -1773142744  1796260105  1463879990   901914175   104491892
## [205]  1605431269 -1933329566  1688405883  -446088064  1238889089   197049934
## [211]  -709469577 -1072333748  1691378909 -1260585478   198644531  2053568216
## [217]   903127801 -1970919834  -473567825  1614821412 -1905604395  1082827666
## [223]  1558537707  1875545136  1518383729 -1265655426 -2085242905  1791098620
## [229]  1447558093 -1153758166   -99557469   -92185464 -2016280343  1847562134
## [235]  1495701791  -221368108   409722309  -429353022  1765302363  2137311200
## [241]  -373658015   273043630  -350916265  -935055956    43404989    52012634
## [247]  1867958291  1488596536 -1347953959   174081222  2002460815  1429165444
## [253]  -205312331  1264621554  -603785525  1270017936 -1543231919 -1282028578
## [259]   908887751   726075484  1269456301 -1680094070  -990917501 -1377014808
## [265] -1279971127  1281050102   228230143  1097770548 -1438663771  1295361058
## [271]   829172027   988808000  1704778305   804328206 -1257113545  -516583668
## [277] -1624037219  1034190522   904064243 -1716316776  1108935353   904106790
## [283]  1222361967  1146561252  1232110741   174767186  2136668075 -1843985680
## [289]   713263665  1133192766  1302119847  -499465796  -425742451  2035727594
## [295]  1324820835  -227988664 -1598926679   227290198   601218783  1836305300
## [301]  1386514821   306372738  -445226469   618852000   -25741791   156697966
## [307]  -345772265 -2126405524  1998516861  -392853734  1588822483  1965665528
## [313] -1658840423 -1901588090  -687876529   -15753148 -1427453323 -1799286606
## [319]   -47880053    97437264  -319365615   688369822  -272731001   469052188
## [325]    27259245  1573117258  -446761405  1976539816  2093047945   424297142
## [331]  1217440191   506831092 -1961736347 -1834464030  1234111227   907381248
## [337]  -247365119   118499278 -1581033993  -893361716 -2100188067   335855482
## [343]    83920563 -1896483752  -323673479  -498745370  2088720687 -2102342236
## [349]  1873412181   226202898 -1483060885  1437743536  -430562831  -190616834
## [355] -1639345305   281953404   857940813  -549769814  -245419229  1375189512
```

```
## [361]  -237346711   590186774    75687071   655107668   151057733   930998594
## [367] -1108466725  1398789472  1995685345  1605663278  1206398167 -1945513172
## [373]  1992513085  1544169434  1610742675  -152048712  -657450407  1247059526
## [379]  1880247311  -124605692   723920437 -1548596878  1827773003   479812880
## [385]   228152785    49698142   922100295 -1524757028  -845069011   534031882
## [391]  -131080189   213485928   636833865   718143350 -1134260353 -2024842316
## [397] -1108831451  1977333154  1053535419  1301926080  -997856831   366738574
## [403] -1450544201  1064694924 -1016336355  -390217670 -1024466829   686789400
## [409] -2056715719   745319590  -999248145 -1240647580 -1395180523 -1837290030
## [415]  -681354453  -514051984  1438153137  2090364862  -209968857  1765574460
## [421]  -544057587  -844603798 -1693909789 -1746073400 -1156960215  2076419542
## [427] -1326601633  1784103188  -683597563  -824593918  1683989915  -509903840
## [433]   183502241  -132206866  -295556457   190629356 -1790739971  1849133210
## [439] -1660799661   214755960 -1837639143   975563526  1750237647  1014527428
## [445]     3490293   552878642   220695563   382907344 -1381266031  1445050910
## [451]  1771278343 -1719553892   862869741   583941834 -1759344189  1365915688
## [457]  -820969463 -1381598154   -19516097   662427252 -1098735899  -812655006
## [463]  1658982011 -1203972224  1999245697 -1592487602 -1708699273 -1038727348
## [469]  -725486627   747602170  2037447219  -161484328   469017081  1897421158
## [475]   644859055   959210276  1824012245 -1573943662  -797561621   466937648
## [481]     6984049  1344943230 -1963692313   507873788  1336756941  -446804182
## [487]  -978024797    50927496   -66994199 -1542552938 -1630130145  1108679636
## [493]   421858501   286669506   176875355  1716904672   841747809  2002101166
## [499] -1936594857  -503678804   643784125  -270685862    -9162989 -1518294728
## [505] -1177069095   450623430 -1518307441 -2055143292  1977097653  1967586034
## [511]  2139569611   993708688   887981393  -146153762 -1521041977 -1948249252
## [517]  1992764589  1735430026   469169027  -492722456  1473540041 -1902921482
## [523]  1705351935  1769673012  -929011035   948225826  -946720709  1824431680
## [529]  1626208577 -1384520178    22671159 -1788782068  -359417955   272236986
## [535]  -230435853  1174868120 -2145910343  -855063002  1748802159   651054564
## [541]  -619908203    89300818   345161387 -1411621392   774662449 -1541883586
## [547]  1651670183   581520572 -1489764723 -2028142614 -1423847325 -1844713912
## [553]  1954615209  -389144746    66876895  2030417556  -361973627  -151813246
## [559] -1573918437   944703904   610784545  1108957294 -1875417577 -1297945748
## [565]  1037500797  1908181530   823650515  1875585016   -22111847  1765196934
## [571]  -849597105  1315720004 -1748059787  -915770446   634433419 -1869504176
## [577]  -887145199  2066662302  -939545721  -822528484 -1687437203 -1367629750
## [583] -1603461821   522180008  1610588041  2052437430   110280895  2014120948
## [589]  -670960027   159018978  1050415611   568272128 -1718509311    -3409202
## [595]   753028343 -1139331892  -123651235 -2072165766 -1222087245   648343384
## [601]  1100161401   486404838   261566511  1504901284  -476745899  1151760402
## [607]  -445050773  -130902864  -423755535  1831075326   934693479   690474876
## [613]  -907644339  -744197974  1158732323    62223624 -1538777239  1455586326
## [619]  -702514273 -1712778924   651699269   959548482  -586241317  1850142816
## [625]  -647799583  2099891502
```

```
# Access accuracy
accuracy3 <- 1 - cv_results2$delta[1]
cat("Accuracy:", accuracy3, "\n")
```

```
## Accuracy: 0.9270332
```

```
# Predict the probabilities of the positive class ("Wins")
predicted_probs3 <- predict(logistic_model2, newdata = lib, type = "response")


# Convert probabilities to binary predictions
predicted_labels3 <- ifelse(predicted_probs3 > 0.1, "readmitted", "Not readmitted")

# Display results
#print(data.frame(Probabilities = predicted_probs3, Predicted_Labels = predicted_labels3))
at <- data.frame(Probabilities = predicted_probs3, Predicted_Labels = predicted_labels3) %>% arrange(desc(Probabilities))
head(at)
```

```
##       Probabilities Predicted_Labels
## 59841     0.1936381       readmitted
## 36935     0.1919764       readmitted
## 38066     0.1919764       readmitted
## 52957     0.1919764       readmitted
## 54845     0.1919764       readmitted
## 98779     0.1919764       readmitted
```

```
# Display the predicted labels
#print(predicted_labels3)
pred_model3 <- lib
pred_model3$predictor <- predicted_labels3
head(pred_model3)
```

```
##       encounter_id patient_nbr      race gender age admission_type_id
## 20447    -1.338575  -0.9452118 Caucasian Female  70       -1.12965500
## 20738    -1.329333  -0.9384905 Caucasian Female  60        0.05693921
## 20825    -1.326727  -0.9376905 Caucasian Female  90       -1.12965500
## 21084    -1.318555  -0.9341307 Caucasian   Male  70       -1.12965500
## 23880    -1.225515  -1.2857602 Caucasian Female  70       -1.12965500
## 23923    -1.224257  -1.0683153 Caucasian   Male  70       -1.12965500
##       admission_type_name discharge_disposition_id
## 20447           Emergency               4.47077231
## 20738              Urgent              -0.42621943
## 20825           Emergency              -0.42621943
## 21084           Emergency              -0.19302935
## 23880           Emergency               0.04016074
## 23923           Emergency               0.04016074
##                                                    discharge_disposition_name
## 20447 Discharged/transferred to another rehab fac including rehab units of a hospital .
## 20738                                                          Discharged to home
## 20825                                                          Discharged to home
## 21084                            Discharged/transferred to another short term hospital
## 23880                                                Discharged/transferred to SNF
## 23923                                                Discharged/transferred to SNF
##       admission_source_id                     admission_source_name
## 20447           0.9104398                             Emergency Room
## 20738          -1.1836534                        Physician Referral
## 20825           0.9104398                             Emergency Room
## 21084           0.9104398                             Emergency Room
## 23880           0.5614243  Transfer from another health care facility
## 23923           0.5614243  Transfer from another health care facility
##       time_in_hospital payer_code          medical_specialty num_lab_procedures
## 20447       0.94629027         MC Orthopedics-Reconstructive          0.8489277
## 20738      -0.41694574         MC                 Nephrology          0.8994109
## 20825      -0.07613673         MC           Emergency/Trauma          0.7479613
## 21084       1.96871727         MC           InternalMedicine          1.3537599
## 23880       2.65033527         UN           InternalMedicine          1.8081089
## 23923       2.65033527         MC           InternalMedicine          0.9498941
##       num_procedures num_medications number_outpatient number_inpatient diag_1
## 20447      0.2383056      -0.1098105         -0.295955       -0.3652328    821
## 20738      0.7901065      -0.5574715         -0.295955       -0.3652328    V56
## 20825     -0.3134954      -0.7813020         -0.295955       -0.3652328    532
## 21084     -0.3134954       0.2259352         -0.295955       -0.3652328    682
## 23880      1.8937085       0.3378505         -0.295955       -0.3652328    238
## 23923      1.8937085      -0.6693867         -0.295955       -0.3652328    532
##       diag_2 diag_3 number_diagnoses max_glu_serum A1Cresult metformin
## 20447    276    285        0.9603645             0         0         0
## 20738    403    599       -0.5801812             0         0         0
## 20825    428    535       -0.5801812             0         0         0
## 21084    427    276       -0.5801812             0         0         1
## 23880  25002    733       -1.0936964             0         8         0
## 23923    280    569       -1.0936964             0         0         0
##       repaglinide nateglinide chlorpropamide glimepiride glipizide glyburide
## 20447           0           0              0           0        No         0
## 20738           0           0              0           0        No         0
## 20825           0           0              0           1        No         0
## 21084           0           0              0           0        No         0
## 23880           0           0              0           0        No         0
## 23923           0           0              0           0        No         0
##       pioglitazone rosiglitazone acarbose miglitol tolazamide insulin
## 20447            2            No        0       No         No       1
## 20738            0            No        0       No         No       1
## 20825            0            No        0       No         No       0
## 21084            0            No        0       No         No       1
## 23880            0            No        0       No         No       1
## 23923            0            No        0       No         No       0
##       glyburide.metformin glipizide.metformin metformin.pioglitazone change
## 20447                  No                   0                     No      1
## 20738                  No                   0                     No      0
## 20825                  No                   0                     No      0
## 21084                  No                   0                     No      1
## 23880                  No                   0                     No      0
## 23923                  No                   0                     No      0
```

```
##       diabetesMed   diagnosis     visits readmitted    num_med num_changes
## 20447           1      Injury -0.4838539          0 -0.1898372  -0.1885303
## 20738           1       Other -0.4838539          0 -0.1898372  -0.1885303
## 20825           1   Digestive -0.4838539          0 -0.1898372  -0.1885303
## 21084           1 Circulatory -0.4838539          0 -0.1898372  -0.1885303
## 23880           1   Neoplasms -0.4838539          0 -0.1898372  -0.1885303
## 23923           0   Digestive -0.4838539          0 -0.1898372  -0.1885303
##            predictor
## 20447 Not readmitted
## 20738 Not readmitted
## 20825 Not readmitted
## 21084 Not readmitted
## 23880 Not readmitted
## 23923 Not readmitted
```

```
# Display the confusion matrix
confusion_matrix3 <- table(predicted_labels3, lib$readmitted)
#print(confusion_matrix3)


# Linear regression
model22 <- lm(as.numeric(as.factor(readmitted)) ~ visits + number_outpatient , data = train_data)
summary(model22)
```

```
##
## Call:
## lm(formula = as.numeric(as.factor(readmitted)) ~ visits + number_outpatient,
##     data = train_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.16422 -0.07548 -0.07548 -0.07548  0.93941
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        1.080710   0.002454 440.350  < 2e-16 ***
## visits             0.019739   0.003285   6.008 1.93e-09 ***
## number_outpatient -0.014609   0.003282  -4.452 8.58e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.272 on 12279 degrees of freedom
## Multiple R-squared:  0.002967,   Adjusted R-squared:  0.002805
## F-statistic: 18.27 on 2 and 12279 DF,  p-value: 1.192e-08
```

```
# Make predictions
predictions5 <- predict(model22, newdata = test_data)

predicted_probs22 <- predict(model22, newdata = test_data , type = "response")

# Convert probabilities to binary predictions
predicted_classes22 <- ifelse(predicted_probs22 > 0.1, "Readmitted", "Not readmitted")
# Calculate RMSE
rmse <- sqrt(mean((as.numeric(as.factor(train_data$readmitted)) - predictions5)^2))
```

```
## Warning in as.numeric(as.factor(train_data$readmitted)) - predictions5: longer
## object length is not a multiple of shorter object length
```

```
# Evaluate accuracy (you might want to use a suitable metric for regression, e.g., RMSE)
accuracy5 <- sqrt(mean((as.numeric(test_data$readmitted) - predictions5)^2))
accuracy5
```

```
## [1] 1.037928
```

```
# Display results
print(paste("Root Mean Squared Error (RMSE):", round(accuracy5, 3)))
```

```
## [1] "Root Mean Squared Error (RMSE): 1.038"
```

```
# Baseline model: Predict the mean of the response variable
baseline_predictions <- rep(mean(as.numeric(train_data$readmitted)), nrow(test_data))

# Calculate RMSE for the baseline model
baseline_rmse <- sqrt(mean((as.numeric(test_data$readmitted) - baseline_predictions)^2))



# Calculate RMSE for the model
model5_rmse <- sqrt(mean((as.numeric(test_data$readmitted) - predictions5)^2))




test_data2 <- test_data
# Compare the RMSE of the baseline model and the new model
print(paste("Baseline RMSE 2:", round(baseline_rmse, 3)))
```

```
## [1] "Baseline RMSE 2: 0.268"
```

```
print(paste("Model 5 RMSE:", round(model5_rmse, 3)))
```

```
## [1] "Model 5 RMSE: 1.038"
```

```
# Make predictions on the test set
test_data2$predicted_readmission <- predict(model22, newdata = test_data2)

head(test_data2)
```

```
##       encounter_id patient_nbr             race gender age admission_type_id
## 24070    -1.219663   -1.100618        Caucasian Female  30       -1.12965500
## 24138    -1.217350   -1.101954        Caucasian Female  70       -1.12965500
## 24277    -1.212543   -1.180311        Caucasian Female  80       -1.12965500
## 24302    -1.211674   -1.144384        Caucasian   Male  30        0.05693921
## 24315    -1.211539   -1.066458 African American Female  80       -1.12965500
## 24345    -1.210605   -1.447905        Caucasian Female  70        0.05693921
##       admission_type_name discharge_disposition_id discharge_disposition_name
## 24070           Emergency               -0.4262194         Discharged to home
## 24138           Emergency               -0.4262194         Discharged to home
## 24277           Emergency               -0.4262194         Discharged to home
## 24302              Urgent                0.9729211                   Left AMA
## 24315           Emergency               -0.4262194         Discharged to home
## 24345              Urgent               -0.4262194         Discharged to home
##       admission_source_id                      admission_source_name
## 24070           0.5614243  Transfer from another health care facility
## 24138           0.9104398                              Emergency Room
## 24277           0.5614243  Transfer from another health care facility
## 24302          -1.1836534                          Physician Referral
## 24315           0.5614243  Transfer from another health care facility
## 24345          -1.1836534                          Physician Referral
##       time_in_hospital payer_code medical_specialty num_lab_procedures
## 24070       2.99114428         MC        Nephrology          1.7071424
## 24138      -0.75775474         MC  InternalMedicine          0.4955452
## 24277      -0.07613673         MC  InternalMedicine          0.1421626
## 24302      -0.75775474         UN  InternalMedicine          0.3440955
## 24315       0.26467227         UN  InternalMedicine          1.2023102
## 24345      -0.75775474         MC  InternalMedicine         -0.9179850
##       num_procedures num_medications number_outpatient number_inpatient diag_1
## 24070      0.2383056     0.002104749         -0.295955       -0.3652328    112
## 24138     -0.8652964    -1.005132450         -0.295955       -0.3652328    577
## 24277     -0.8652964    -0.557471472         -0.295955       -0.3652328    414
## 24302     -0.8652964    -1.117047694         -0.295955       -0.3652328    486
## 24315     -0.8652964    -1.005132450         -0.295955       -0.3652328    435
## 24345     -0.8652964    -0.781301961         -0.295955       -0.3652328    428
##       diag_2 diag_3 number_diagnoses max_glu_serum A1Cresult metformin
## 24070    996  25013        -1.093696             0         8         0
## 24138    428    414        -1.093696             0         0         0
## 24277    427    424        -1.093696             0         0         0
## 24302    493  25013        -1.093696             0         0         0
## 24315    427  25092        -1.093696             0         0         1
## 24345    250    414        -1.093696             0         0         0
##       repaglinide nateglinide chlorpropamide glimepiride glipizide glyburide
## 24070           0           0              0           0        No         0
## 24138           0           0              0           0        No         0
## 24277           0           0              0           1        No         0
## 24302           0           0              0           0        No         0
## 24315           0           0              0           0    Steady         0
## 24345           0           0              0           0        No         0
##       pioglitazone rosiglitazone acarbose miglitol tolazamide insulin
## 24070            0            No        0       No         No       1
## 24138            0            No        0       No         No       0
## 24277            0            No        0       No         No       1
## 24302            0            No        0       No         No       0
## 24315            0            No        0       No         No       0
## 24345            0            No        0       No         No       1
##       glyburide.metformin glipizide.metformin metformin.pioglitazone change
## 24070                  No                   0                      No      0
## 24138                  No                   0                      No      0
## 24277                  No                   0                      No      1
## 24302                  No                   0                      No      0
## 24315                  No                   0                      No      1
## 24345                  No                   0                      No      0
##       diabetesMed   diagnosis     visits readmitted    num_med num_changes
## 24070           1 Circulatory -0.4838539          0 -0.1898372  -0.1885303
## 24138           0   Digestive -0.4838539          0 -0.1898372  -0.1885303
## 24277           1 Circulatory -0.4838539          0 -0.1898372  -0.1885303
## 24302           0 Respiratory -0.4838539          0 -0.1898372  -0.1885303
## 24315           1 Circulatory -0.4838539          0 -0.1898372  -0.1885303
## 24345           1 Circulatory -0.4838539          0 -0.1898372  -0.1885303
```

```
##       predicted_readmission
## 24070              1.075483
## 24138              1.075483
## 24277              1.075483
## 24302              1.075483
## 24315              1.075483
## 24345              1.075483
```

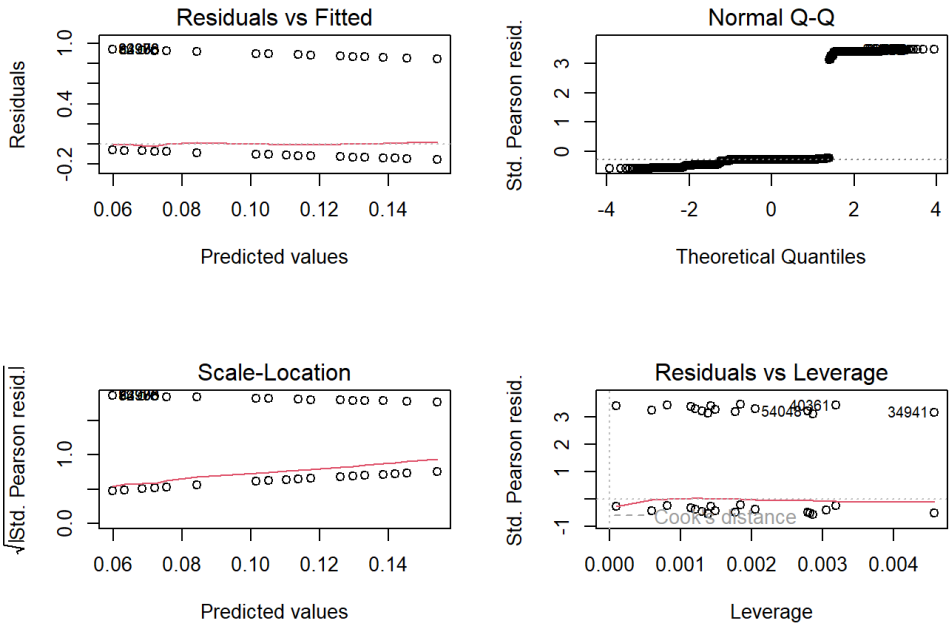
```
test_data_final <- test_data2 %>% arrange(desc(predicted_readmission))

test_data_final$predicted_readmission <- round(test_data_final$predicted_readmission,0)

##test_data_final %>% arrange(predicted_readmission)
accuracy_table <- data.frame(Mult_regress_1 = round(accuracy,3),
                             Mult_regress_2 = round(accuracy2,3),
                             Mult_regress_3 = round(accuracy3,3),
                             log_regress = round(accuracy5,3),
                             baseline_rmse_log = round(baseline_rmse,3))

confusion_matrix22 <- table(predicted_classes22, test_data$readmitted)
print(confusion_matrix22)
```

```
##
## predicted_classes22    0    1
##          Readmitted 4855  410
```

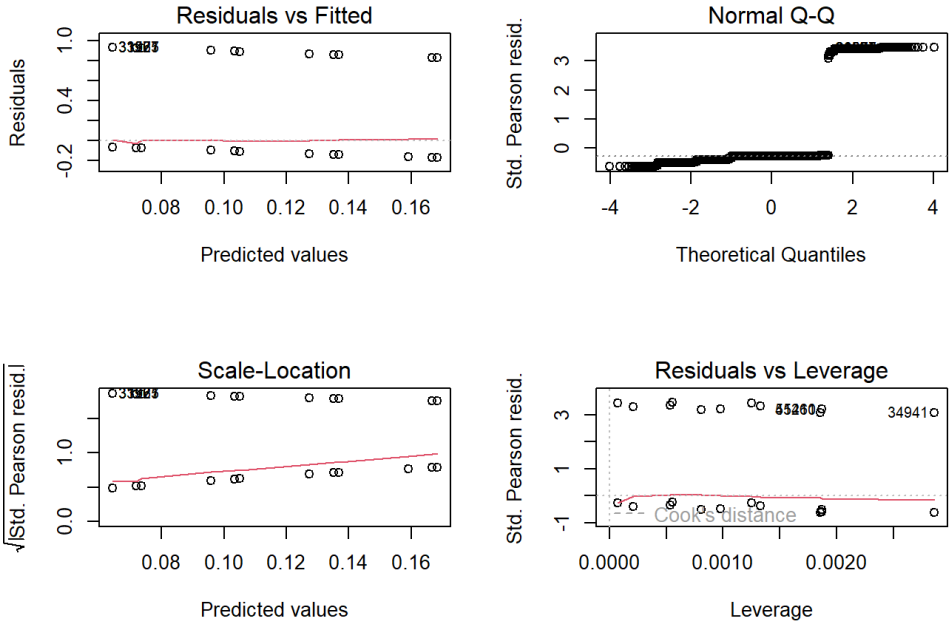
```
htmlTable( accuracy_table)
```

	Mult_regress_1	Mult_regress_2	Mult_regress_3	log_regress	baseline_rmse_log
1	0.825	0.927	0.927	1.038	0.268

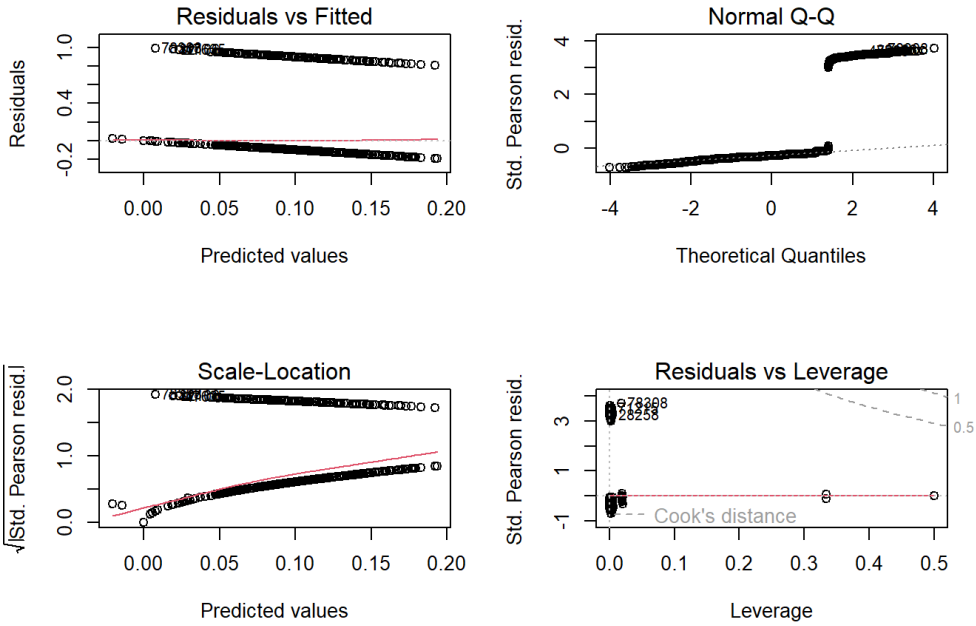
```
# Plot the residuals
par(mfrow=c(2,2))
plot(model)
```



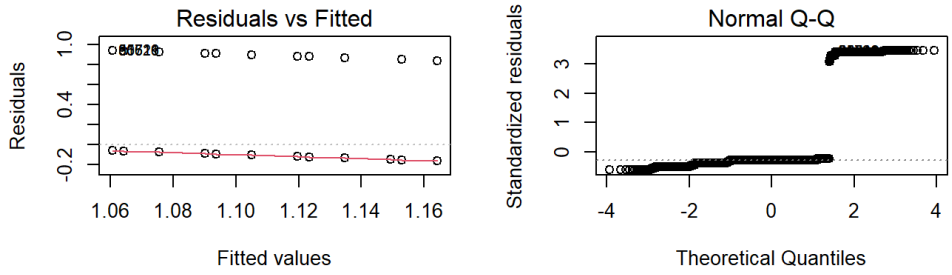

```
par(mfrow=c(2,2))
plot(logistic_model)
```



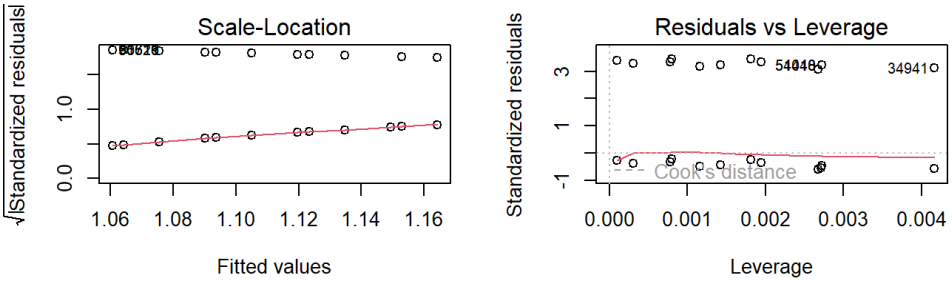

```
par(mfrow=c(2,2))
plot(logistic_model2)
```




```
par(mfrow=c(2,2))
plot(model22)
```



Conclusion reducing readmission



today is one of the major goals hospital in the USA has. Hospital can create model that can predict readmission on with 30 days of the first visit. the difficulty of the creating a model will require data processing skills and statistics. we were able to predict 2529 readmission from patient that weren't readmitted while predict 314 from the ones that were predicted.