admissions <- read.csv(file = "ADMISSIONS.csv")

caregivers <- read.csv(file = "CAREGIVERS.csv")

callout <- read.csv(file = "CALLOUT.csv")

patients <- read.csv(file ="PATIENTS.csv")

icustays <- read.csv(file = "ICUSTAYS.csv")

chartevents <- read.csv(file = "CHARTEVENTS.csv")

d\_items <- read.csv(file = "D\_ITEMS.csv")

labevents <- read.csv( file = "LABEVENTS.csv")

d\_labitems <- read.csv("D\_LABITEMS.csv")

> str(icustays)

'data.frame': 136 obs. of 12 variables:

$ row\_id : int 12742 12747 12749 12754 12755 12762 12763 12765 12767 12768 ...

$ subject\_id : int 10006 10011 10013 10017 10019 10026 10027 10029 10032 10033 ...

$ hadm\_id : int 142345 105331 165520 199207 177759 103770 199395 132349 140372 157235 ...

$ icustay\_id : int 206504 232110 264446 204881 228977 277021 286020 226055 267090 254543 ...

$ dbsource : chr "carevue" "carevue" "carevue" "carevue" ...

$ first\_careunit: chr "MICU" "MICU" "MICU" "CCU" ...

$ last\_careunit : chr "MICU" "MICU" "MICU" "CCU" ...

$ first\_wardid : int 52 15 15 7 15 33 12 33 52 33 ...

$ last\_wardid : int 52 15 15 7 15 33 12 33 52 33 ...

$ intime : chr "10/23/2164 21:10" "8/14/2126 22:34" "10/4/2125 23:38" "5/29/2149 18:52" ...

$ outtime : chr "10/25/2164 12:21" "8/28/2126 18:59" "10/7/2125 15:13" "5/31/2149 22:19" ...

$ los : num 1.63 13.85 2.65 2.14 1.29 ...

> first\_ICU\_stay <- icustays %>%

+ left\_join(admissions, by = c("subject\_id" = "subject\_id", "hadm\_id" = "hadm\_id")) %>%

+ group\_by(subject\_id) %>%

+ filter(intime == min(intime)) %>%

+ ungroup() %>%

+ select(subject\_id, hadm\_id, icustay\_id, intime, los, diagnosis, hospital\_expire\_flag)

> last\_ICU\_stay <- icustays %>%

+ left\_join(admissions, by = c("subject\_id" = "subject\_id", "hadm\_id" = "hadm\_id")) %>%

+ group\_by(subject\_id) %>%

+ filter(intime == max(intime)) %>%

+ ungroup() %>%

+ select(subject\_id, hadm\_id, icustay\_id, intime,los, diagnosis, hospital\_expire\_flag)

> t1 <- first\_ICU\_stay %>% filter(hospital\_expire\_flag == 1) %>% count()

> t2 <- last\_ICU\_stay %>% filter(hospital\_expire\_flag == 1) %>% count()

> kable(list(t1, t2)) %>% kable\_styling()

> str(chartevents)

'data.frame': 758355 obs. of 15 variables:

$ row\_id : int 5279021 5279022 5279023 5279024 5279025 5279026 5279027 5279028 5279029 5279030 ...

$ subject\_id : int 40124 40124 40124 40124 40124 40124 40124 40124 40124 40124 ...

$ hadm\_id : int 126179 126179 126179 126179 126179 126179 126179 126179 126179 126179 ...

$ icustay\_id : int 279554 279554 279554 279554 279554 279554 279554 279554 279554 279554 ...

$ itemid : int 223761 224695 220210 220045 220179 220180 220181 220046 220047 223751 ...

$ charttime : chr "2/4/2130 4:00" "2/4/2130 4:25" "2/4/2130 4:30" "2/4/2130 4:32" ...

$ storetime : chr "2/4/2130 4:35" "2/4/2130 5:55" "2/4/2130 4:43" "2/4/2130 4:43" ...

$ cgid : int 19085 18999 21452 21452 21452 21452 21452 19085 19085 19085 ...

$ value : chr "95.9" "2222221.7" "15" "94" ...

$ valuenum : num 9.59e+01 2.22e+06 1.50e+01 9.40e+01 1.63e+02 ...

$ valueuom : chr "?F" "cmH2O" "insp/min" "bpm" ...

$ warning : int 0 0 0 0 0 0 0 0 0 0 ...

$ error : int 0 0 0 0 0 0 0 0 0 0 ...

$ resultstatus: chr "" "" "" "" ...

$ stopped : chr "" "" "" "" ...

> head(chartevents)

row\_id subject\_id hadm\_id icustay\_id itemid charttime storetime cgid value valuenum

1 5279021 40124 126179 279554 223761 2/4/2130 4:00 2/4/2130 4:35 19085 95.9 95.9

2 5279022 40124 126179 279554 224695 2/4/2130 4:25 2/4/2130 5:55 18999 2222221.7 2222221.7

3 5279023 40124 126179 279554 220210 2/4/2130 4:30 2/4/2130 4:43 21452 15 15.0

4 5279024 40124 126179 279554 220045 2/4/2130 4:32 2/4/2130 4:43 21452 94 94.0

5 5279025 40124 126179 279554 220179 2/4/2130 4:32 2/4/2130 4:43 21452 163 163.0

6 5279026 40124 126179 279554 220180 2/4/2130 4:32 2/4/2130 4:43 21452 81 81.0

valueuom warning error resultstatus stopped

1 ?F 0 0

2 cmH2O 0 0

3 insp/min 0 0

4 bpm 0 0

5 mmHg 0 0

6 mmHg 0 0

> str(labevents)

'data.frame': 76074 obs. of 9 variables:

$ row\_id : int 6244563 6244564 6244565 6244566 6244567 6244568 6244569 6244570 6244571 6244572 ...

$ subject\_id: int 10006 10006 10006 10006 10006 10006 10006 10006 10006 10006 ...

$ hadm\_id : int NA NA NA NA NA NA NA NA NA NA ...

$ itemid : int 50868 50882 50893 50902 50912 50931 50960 50970 50971 50983 ...

$ charttime : chr "9/24/2164 20:21" "9/24/2164 20:21" "9/24/2164 20:21" "9/24/2164 20:21" ...

$ value : chr "19" "27" "10" "97" ...

$ valuenum : num 19 27 10 97 7 126 2.3 5.6 4.3 139 ...

$ valueuom : chr "mEq/L" "mEq/L" "mg/dL" "mEq/L" ...

$ flag : chr "" "" "" "" ...

> last\_ICU\_stay <- last\_ICU\_stay %>% mutate(diagnosis\_type = case\_when(

+ grepl("(.)\*CANCER|(.) CA|(.)\*LEUKEMIA", diagnosis, ignore.case = TRUE) ~ "CANCER",

+ grepl("(.)\*SEPSIS|INFECTION|CHOLANGITIS|ABSCESS|FEVER|PNEUMONIA", diagnosis, ignore.case = TRUE) ~ "INFECTION",

+ grepl("(.)\*MI|VF ARREST|STROKE(.)\*|CHEST PAIN|MYOCARDIAL", diagnosis, ignore.case = TRUE) ~ "ACUTE CARDIAC",

+ grepl("(.)\*ACCIDENT|FRACTURE|(.)\*FALL|SEIZURE|BREATH|BLEED|HEMATOMA|OVERDOSE|SYNCOPE(.)\*|TACHYPNEA|ACUTE",diagnosis, ignore.case = TRUE) ~ "ACUTE",

+ grepl("(.)\*PULMONARY|RESPIRATORY|LUNG|ASTHMA|LIVER|HEPATI(TIS|IC)|ESOPHAGEAL|OA|ARTHRITIS|(HYPO|HYPER)TENSION|HEADACHE|DISTRESS|MENTAL|HEAD|ANGINA|BRAIN|HEART FAILURE|(.)\*EFFUSION|FAILURE TO THRIVE|"

+ , diagnosis, ignore.case = TRUE) ~ "CHRONIC", TRUE ~ "OTHERS"))

> ggplot(last\_ICU\_stay, aes(x=diagnosis\_type, fill = factor(hospital\_expire\_flag)))+

+ geom\_bar(stat = "count", position = "dodge")+

+ theme\_bw()+

+ theme(axis.text.x = element\_text(size = 8, angle = 90))+

+ labs(x= "Diagnosis Type", y="Case Counts", title = "Diagnosis Type Case Counts")+

+ scale\_y\_continuous(breaks = c(0,1,2,3,4,5,8,12))+

+ scale\_fill\_discrete(name="Death", labels=c("0 (n=85)","1 (n=100)"))

> mosaicplot(table(last\_ICU\_stay$diagnosis\_type, last\_ICU\_stay$hospital\_expire\_flag),

+ main = "Diagnosis Type by Death 1", shade = TRUE, las = 2)

> d\_items %>%

+ filter(str\_detect(label, pattern = regex("heart rate", ignore\_case = TRUE))) %>%

+ print()

row\_id itemid label abbreviation dbsource linksto category

1 212 211 Heart Rate carevue chartevents

2 1078 3494 Lowest Heart Rate carevue chartevents

3 12712 220045 Heart Rate HR metavision chartevents Routine Vital Signs

4 12713 220046 Heart rate Alarm - High HR Alarm - High metavision chartevents Alarms

5 12714 220047 Heart Rate Alarm - Low HR Alarm - Low metavision chartevents Alarms

unitname param\_type conceptid

1 NA

2 NA

3 bpm Numeric NA

4 bpm Numeric NA

5 bpm Numeric NA

> heart\_rate <- chartevents %>%

+ filter(itemid %in% c(211, 220045)) %>%

+ select(subject\_id, hadm\_id, icustay\_id, charttime, valuenum) %>%

+ group\_by(subject\_id, hadm\_id, icustay\_id) %>%

+ summarise(max\_heart\_rate = max(valuenum, na.rm = TRUE))

`summarise()` has grouped output by 'subject\_id', 'hadm\_id'. You can override using the `.groups` argument.

> #Lowest blood pressure

> d\_items %>%

+ filter(str\_detect(label, pattern = regex("diastolic(blood pressure)?|NBP|Arterial BP", ignore\_case = TRUE))) %>%

+ print()

> blood\_pressure <- chartevents %>%

+ filter(itemid %in% c(227242, 224643, 220180, 220051,455, 8441, 5817, 8551, 8368)) %>%

+ select(subject\_id, hadm\_id, icustay\_id, charttime, valuenum) %>%

+ group\_by(subject\_id, hadm\_id, icustay\_id) %>%

+ summarise(min\_blood\_pressure = min(valuenum, na.rm = TRUE))

` summarise()` has grouped output by 'subject\_id', 'hadm\_id'. You can override using the `.groups` argument.

> last\_ICU\_stay <- last\_ICU\_stay %>%

+ left\_join(heart\_rate, by = c("subject\_id", "hadm\_id", "icustay\_id")) %>%

+ left\_join(blood\_pressure, by = c("subject\_id", "hadm\_id", "icustay\_id"))

> last\_ICU\_stay %>% filter(is.na(last\_ICU\_stay$max\_heart\_rate))

# A tibble: 2 x 11

subject\_id hadm\_id icustay\_id intime los diagnosis hospital\_expire\_~ diagnosis\_type min\_GCS

<int> <int> <int> <chr> <dbl> <chr> <int> <chr> <dbl>

1 10067 160442 236674 10/6/2130 ~ 0.106 S/P MOTORCYCLE ~ 1 ACUTE Inf

2 10120 193924 268282 5/12/2115 ~ 2.65 LIVER FAILURE 1 CHRONIC Inf

# ... with 2 more variables: max\_heart\_rate <dbl>, min\_blood\_pressure <dbl>

> ggplot(last\_ICU\_stay, aes(x= as.character(hospital\_expire\_flag),

+ y= max\_heart\_rate,

+ fill = factor(hospital\_expire\_flag)))+

+ geom\_boxplot(position = "dodge")+ theme\_bw()+

+ labs(x= "Death", y="Heart Rate(BPM)",

+ title = "Maximum Heart Rate During the ICU Stay")+

+ scale\_fill\_discrete(name="Death", labels=c("0 (n=100)","1 (n=120)"))

Warning message:

Removed 2 rows containing non-finite values (stat\_boxplot).

> ggplot(last\_ICU\_stay, aes(x= as.character(hospital\_expire\_flag),

+ y= min\_blood\_pressure,

+ fill = factor(hospital\_expire\_flag)))+

+ geom\_boxplot(position = "dodge")+ theme\_bw()+

+ labs(x= "Death", y="Blood Pressure(mmHg)",

+ title = "Lowest Blood Pressure During the ICU Stay")+

+ scale\_fill\_discrete(name="Death", labels=c("0 (n=100)","1 (n=120)"))

Warning message:

Removed 2 rows containing non-finite values (stat\_boxplot).

> d\_labitems %>%

+ filter(str\_detect(label, pattern = regex("Glucose", ignore\_case = TRUE))) %>%

+ print()

row\_id itemid label fluid category loinc\_code

1 10 50809 Glucose Blood Blood Gas 2339-0

2 43 50842 Glucose, Ascites Ascites Chemistry 2347-3

3 132 50931 Glucose Blood Chemistry 2345-7

4 214 51014 Glucose, CSF Cerebrospinal Fluid (CSF) Chemistry 2342-4

5 222 51022 Glucose, Joint Fluid Joint Fluid Chemistry 2348-1

6 234 51034 Glucose, Body Fluid Other Body Fluid Chemistry 2344-0

7 253 51053 Glucose, Pleural Pleural Chemistry 2346-5

8 284 51084 Glucose, Urine Urine Chemistry 2350-7

9 678 51478 Glucose Urine Hematology 5792-7

10 729 51529 Estimated Actual Glucose BLOOD CHEMISTRY

> min\_BG<- labevents %>%

+ filter(itemid %in% c(50809, 50931)) %>%

+ group\_by(subject\_id, hadm\_id) %>%

+ summarise(min\_BG = min(valuenum, na.rm = T))

`summarise()` has grouped output by 'subject\_id'. You can override using the `.groups` argument.

> last\_ICU\_stay <- last\_ICU\_stay %>%

+ left\_join(min\_BG, by=c("subject\_id", "hadm\_id"))

> #the length of ICU stay

> last\_ICU\_stay %>% group\_by(hospital\_expire\_flag) %>%

+ summarise(mean=mean(los))

# A tibble: 2 x 2

hospital\_expire\_flag mean

<int> <dbl>

1 0 4.09

2 1 5.82

> ggplot(last\_ICU\_stay, aes(x= as.character(hospital\_expire\_flag),

+ y= min\_BG,

+ fill = factor(hospital\_expire\_flag)))+

+ geom\_boxplot(position = "dodge")+ theme\_bw()+

+ labs(x= "Death", y="Blood Glucose(mg/dL)",

+ title = "Lowest Blood Glucose During the Hospital Admission")+

+ scale\_fill\_discrete(name="Death", labels=c("0 (n=100)","1 (n=140)"))

> ggplot(last\_ICU\_stay, aes(x= as.character(hospital\_expire\_flag),

+ y= los,

+ fill = factor(hospital\_expire\_flag)))+

+ geom\_boxplot(position = "dodge")+ theme\_bw()+

+ labs(x= "Death", y="Days",

+ title = "The Length of ICU Stay")+

+ scale\_fill\_discrete(name="Death", labels=c("0 (n=100)","1 (n=140)"))+

+ annotate("text", x=1, y=40, label="average 3.85 days in ICU ")+

+ annotate("text", x=2, y=40, label="average 6.19 days in ICU ")

> library(lubridate)

> gender <- patients %>%

+ select(subject\_id, gender) %>%

+ mutate(male = case\_when(gender == "M" ~1,

+ TRUE ~0)) %>%

+ right\_join(last\_ICU\_stay) %>%

+ select(subject\_id, hadm\_id, icustay\_id, male, hospital\_expire\_flag)

> date\_of\_birth <- patients %>%

+ select(subject\_id, dob) %>%

+ right\_join(last\_ICU\_stay) %>%

+ mutate(

+ dob = mdy\_hm(dob),

+ intime = mdy\_hm(intime),

+ age\_intake\_ICU = round(as.numeric(intime - dob)/365.25)

+ )%>%

+ select(subject\_id, hadm\_id, icustay\_id, age\_intake\_ICU, hospital\_expire\_flag)

> table(is.na(date\_of\_birth$age\_intake\_ICU))

FALSE TRUE

92 8

> last\_ICU\_stay <- last\_ICU\_stay %>%

+ left\_join(gender) %>%

+ left\_join(date\_of\_birth)

> analysis\_data <- last\_ICU\_stay

> factor\_vars <- c('los','diagnosis\_type', "min\_GCS", "max\_heart\_rate"

+ , "min\_blood\_pressure", "min\_BG", "male")

> last\_ICU\_stay[factor\_vars] <- lapply(last\_ICU\_stay[factor\_vars], function(x) as.factor(x))

> set.seed(20210116)

> library(mice)

> mice\_mod <- mice(last\_ICU\_stay[, !names(last\_ICU\_stay) %in% c('subject\_id',

+ 'hadm\_id', 'icustay\_id', 'intime', 'hospital\_expire\_flag')], method='rf')

> mice\_output <- complete(mice\_mod)

> par(mfrow=c(1,2))

> hist(last\_ICU\_stay$age\_intake\_ICU, freq = F, main = "Age: Original Data",

+ col = 'darkgreen', ylim = c(0, 0.04))

> hist(mice\_output$age\_intake\_ICU, freq = F, main = "Age: MICE Output",

+ col = 'lightgreen', ylim = c(0, 0.04))

>

> mice\_output <- complete(mice\_mod)

> library(rsample)

> set.seed(2021)

> data\_split <- initial\_split(analysis\_data, prop = 7/10)

> training\_data <- training(data\_split)

> testing\_data <- testing(data\_split)

> model\_bin <- training\_data %>%

+ glm(formula = hospital\_expire\_flag ~ los+diagnosis\_type

+ +max\_heart\_rate+min\_blood\_pressure+male+age\_intake\_ICU,

+ family = 'binomial',na.action = na.omit)

> summary(model\_bin)

Call:

glm(formula = hospital\_expire\_flag ~ los + diagnosis\_type + max\_heart\_rate +

min\_blood\_pressure + male + age\_intake\_ICU, family = "binomial",

data = ., na.action = na.omit)

Deviance Residuals:

Min 1Q Median 3Q Max

-2.3928 -0.8309 -0.4283 1.0465 2.1282

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -3.45443 3.39163 -1.019 0.3084

los 0.03528 0.06395 0.552 0.5812

diagnosis\_typeACUTE CARDIAC -0.48649 0.92732 -0.525 0.5998

diagnosis\_typeCANCER -3.17787 1.53799 -2.066 0.0388 \*

diagnosis\_typeCHRONIC -1.39977 0.82103 -1.705 0.0882 .

diagnosis\_typeINFECTION -1.81710 0.94747 -1.918 0.0551 .

max\_heart\_rate 0.02357 0.01804 1.306 0.1915

min\_blood\_pressure -0.04106 0.02459 -1.670 0.0949 .

male 0.46198 0.66267 0.697 0.4857

age\_intake\_ICU 0.03278 0.02472 1.326 0.1847

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 91.422 on 68 degrees of freedom

Residual deviance: 71.381 on 59 degrees of freedom

AIC: 91.381

Number of Fisher Scoring iterations: 5

> training\_roc <- training\_data %>%

+ ggplot(aes(m= predicted\_outcome, d=hospital\_expire\_flag))+

+ labs(title = "ROC Plot for Training Data")+

+ geom\_roc(n.cuts = 10, labels = F, labelround = 4)+

+ style\_roc(theme = theme\_grey)

> calc\_auc(training\_roc)$AUC\*100

[1] 100

> training\_data$predicted\_outcome <- predict(model\_bin, training\_data,

+ type = "response")

> training\_roc <- training\_data %>%

+ ggplot(aes(m= predicted\_outcome, d=hospital\_expire\_flag))+

+ labs(title = "ROC Plot for Training Data")+

+ geom\_roc(n.cuts = 10, labels = F, labelround = 4)+

+ style\_roc(theme = theme\_grey)

> calc\_auc(training\_roc)$AUC\*100

[1] 84.07871

> testing\_data$predicted\_outcome <- predict(model\_bin, testing\_data,

+ type = "response")

> testing\_roc <- testing\_data %>%

+ ggplot(aes(m= predicted\_outcome, d= hospital\_expire\_flag))+

+ labs(title = "ROC Plot for Testing Data")+

+ geom\_roc(n.cuts = 10, labels = F, labelround = 4) +

+ style\_roc(theme = theme\_grey)

> calc\_auc(testing\_roc)$AUC\*100

[1] 55.95238

> model\_rf <- randomForest(factor(hospital\_expire\_flag) ~ los+diagnosis\_type

+ +max\_heart\_rate

+ +min\_blood\_pressure+male+age\_intake\_ICU,

+ data = analysis\_data, na.action = na.exclude)

> print(model\_rf)

Call:

randomForest(formula = factor(hospital\_expire\_flag) ~ los + diagnosis\_type + max\_heart\_rate + min\_blood\_pressure + male + age\_intake\_ICU, data = analysis\_data, na.action = na.exclude)

Type of random forest: classification

Number of trees: 500

No. of variables tried at each split: 2

OOB estimate of error rate: 35.71%

Confusion matrix:

0 1 class.error

0 51 13 0.2031250

1 22 12 0.6470588

> par(mfrow=c(1,1))

Warning message:

In diff.default(xscale) : reached elapsed time limit

> plot(model\_rf, ylim=c(0,0.55))

> legend("bottomleft", colnames(model\_rf$err.rate), col = 1:3, fill = 1:3)

> library(ggplot2)

> importance <- importance(model\_rf)

> varImprtance <- data.frame(variables = row.names(importance),

+ Importance = round(importance[, 'MeanDecreaseGini'], 2))

> rankImportance <- varImprtance %>%

+ mutate(Rank = paste0('#', dense\_rank(desc(Importance))))

>

> ggplot(rankImportance, aes(x= reorder(variables, Importance),

+ y = Importance, fill= Importance))+

+ geom\_bar(stat = 'identity') +

+ geom\_text(aes(x= variables, y=0.5, label = Rank),

+ hjust=0, vjust=0.55, size=4, color='red')+

+ labs(x='Variables')+

+ coord\_flip()+

+ theme\_bw()