Project 2: Building a Safer Health System

if(!require(pacman)) install.packages(pacman)

## Loading required package: pacman

library(pacman)  
pacman::p\_load(tidyverse, psych, readxl, dplyr, rpart, rattle)

#Plugs file into variable  
ICUEvents <- read\_excel("ICUEvents.xlsx")

## New names:  
## \* `` -> ...11

names(ICUEvents)

## [1] "Report-number" "Facility-ID"   
## [3] "Error-category" "Date-of-error"   
## [5] "Time-of-error" "Day-of-week"   
## [7] "Type-of-error" "Cause-of-error"   
## [9] "Contributing-factor" "Location-of-error"   
## [11] "...11" "Description-of-error"   
## [13] "Medication-process-node" "Staff-type-initiated-error"   
## [15] "Staff-type-perpetuated-error" "Staff-type-discovered-error"   
## [17] "Action-taken" "Action-taken-detail"   
## [19] "Date-record-was-entered"

newICU1 <- ICUEvents  
newICU1 <- newICU1[complete.cases(newICU1["Type-of-error"]), ]

# Remove useless column  
newICU1 <- subset(newICU1, select= -...11)  
newICU2 <- ICUEvents

# Lets you view document  
view(newICU1)

#Summarizes doc  
summary(newICU1)

## Report-number Facility-ID Error-category Date-of-error   
## Min. : 9475 Min. :11466296 Length:30608 Length:30608   
## 1st Qu.: 14149 1st Qu.:35414425 Class :character Class :character   
## Median : 16880 Median :54716378 Mode :character Mode :character   
## Mean : 20079 Mean :55428383   
## 3rd Qu.: 20366 3rd Qu.:74455222   
## Max. :249347 Max. :99276649   
## Time-of-error Day-of-week Type-of-error Cause-of-error   
## Length:30608 Length:30608 Length:30608 Length:30608   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
## Contributing-factor Location-of-error Description-of-error  
## Length:30608 Length:30608 Length:30608   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character   
##   
##   
##   
## Medication-process-node Staff-type-initiated-error  
## Length:30608 Length:30608   
## Class :character Class :character   
## Mode :character Mode :character   
##   
##   
##   
## Staff-type-perpetuated-error Staff-type-discovered-error Action-taken   
## Length:30608 Length:30608 Length:30608   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character   
##   
##   
##   
## Action-taken-detail Date-record-was-entered  
## Length:30608 Length:30608   
## Class :character Class :character   
## Mode :character Mode :character   
##   
##   
##

#Gives you the data types of features  
str(newICU1)

## tibble[,18] [30,608 × 18] (S3: tbl\_df/tbl/data.frame)  
## $ Report-number : num [1:30608] 9475 9475 9477 9477 9498 ...  
## $ Facility-ID : num [1:30608] 71778745 71778745 71778745 71778745 33299661 ...  
## $ Error-category : chr [1:30608] "C" "C" "B" "B" ...  
## $ Date-of-error : chr [1:30608] "01/03/2000" "01/03/2000" "01/03/2000" "01/03/2000" ...  
## $ Time-of-error : chr [1:30608] "00:00" "00:00" "14:00" "14:00" ...  
## $ Day-of-week : chr [1:30608] "Monday" "Monday" "Monday" "Monday" ...  
## $ Type-of-error : chr [1:30608] "Omission error" "Omission error" "Omission error" "Omission error" ...  
## $ Cause-of-error : chr [1:30608] "Contraindicated, drug allergy" "Documentation" "Contraindicated, drug allergy" "Documentation" ...  
## $ Contributing-factor : chr [1:30608] NA NA NA NA ...  
## $ Location-of-error : chr [1:30608] "Nursing (Patient Care) Unit" "Nursing (Patient Care) Unit" "Nursing (Patient Care) Unit" "Nursing (Patient Care) Unit" ...  
## $ Description-of-error : chr [1:30608] "Allergy missing on doctor's order" "Allergy missing on doctor's order" "Allegy missing on doctor's orders" "Allegy missing on doctor's orders" ...  
## $ Medication-process-node : chr [1:30608] "Prescribing" "Prescribing" "Documenting" "Documenting" ...  
## $ Staff-type-initiated-error : chr [1:30608] "Physician" "Physician" "Physician" "Physician" ...  
## $ Staff-type-perpetuated-error: chr [1:30608] NA NA NA NA ...  
## $ Staff-type-discovered-error : chr [1:30608] NA NA NA NA ...  
## $ Action-taken : chr [1:30608] NA NA NA NA ...  
## $ Action-taken-detail : chr [1:30608] NA NA NA NA ...  
## $ Date-record-was-entered : chr [1:30608] "01/04/2000" "01/04/2000" "01/04/2000" "01/04/2000" ...

# Factor   
newICU1$`Date-of-error` <- as.Date(newICU1$`Date-of-error`, tryFormats = c("%d-%m-%Y", "%d/%m/%y"))  
#newICU1$`Time-of-error` <- as.Date(newICU1$`Time-of-error`, tryFormats = c("%h:%m"))  
#newICU1  
#newICU1$`Staff-type-initiated-error`[newICU1$`Staff-type-initiated-error`.con]

newICU1$`Error-category` <- factor(newICU1$`Error-category`)

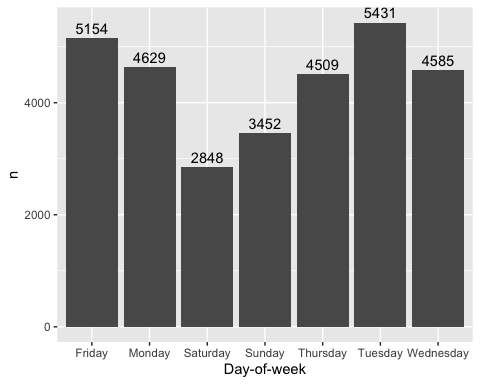
newICU1$`Type-of-error` <- factor(newICU1$`Type-of-error`)  
newICU1$`Cause-of-error` <- factor(newICU1$`Cause-of-error`)

newICU1$`Contributing-factor` <- factor(newICU1$`Contributing-factor`)  
newICU1$`Medication-process-node` <- factor(newICU1$`Medication-process-node`)  
newICU1$`Staff-type-initiated-error` <- factor(newICU1$`Staff-type-initiated-error`)  
newICU1$`Staff-type-perpetuated-error` <- factor(newICU1$`Staff-type-perpetuated-error`)  
newICU1$`Action-taken` <- factor(newICU1$`Action-taken`)  
newICU1$`Staff-type-discovered-error` <- factor(newICU1$`Staff-type-discovered-error`)  
newICU1$`Day-of-week` <- factor(newICU1$`Day-of-week`)

summary(newICU1)

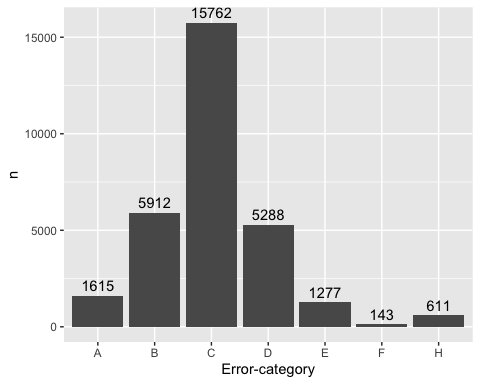
## Report-number Facility-ID Error-category Date-of-error   
## Min. : 9475 Min. :11466296 A: 1615 Min. :2020-01-01   
## 1st Qu.: 14149 1st Qu.:35414425 B: 5912 1st Qu.:2020-03-01   
## Median : 16880 Median :54716378 C:15762 Median :2020-06-02   
## Mean : 20079 Mean :55428383 D: 5288 Mean :2020-05-31   
## 3rd Qu.: 20366 3rd Qu.:74455222 E: 1277 3rd Qu.:2020-09-02   
## Max. :249347 Max. :99276649 F: 143 Max. :2020-12-03   
## H: 611 NA's :17580   
## Time-of-error Day-of-week Type-of-error   
## Length:30608 Friday :5154 Omission error :7404   
## Class :character Monday :4629 Improper dose/quantity:6319   
## Mode :character Saturday :2848 Unauthorized drug :3752   
## Sunday :3452 Prescribing error :2255   
## Thursday :4509 Extra dose :2014   
## Tuesday :5431 Wrong drug preparation:1847   
## Wednesday:4585 (Other) :7017   
## Cause-of-error Contributing-factor  
## Performance (human) deficit : 7218 Distractions : 5125   
## Procedure/protocol not followed : 3627 Workload increase : 1834   
## Transcription inaccurate/omitted: 1763 Staff, inexperienced : 1507   
## Communication : 1723 Staffing, insufficient: 970   
## Knowledge deficit : 1702 Shift change : 946   
## (Other) :12507 (Other) : 5279   
## NA's : 2068 NA's :14947   
## Location-of-error Description-of-error Medication-process-node  
## Length:30608 Length:30608 Administering :11677   
## Class :character Class :character Dispensing : 5690   
## Mode :character Mode :character Documenting : 7134   
## Does not apply: 1615   
## Monitoring : 338   
## Prescribing : 4154   
##   
## Staff-type-initiated-error Staff-type-perpetuated-error  
## Nurse, Registered :12252 Nurse, Registered : 7950   
## Pharmacist : 4412 Pharmacist : 2474   
## Physician : 2902 Pharmacy Technician : 803   
## Pharmacy Technician : 1644 Physician : 747   
## Unit Secretary/Clerk: 1445 Unit Secretary/Clerk: 707   
## (Other) : 5677 (Other) : 2579   
## NA's : 2276 NA's :15348   
## Staff-type-discovered-error  
## Nurse, Registered :15569   
## Pharmacist : 4358   
## Physician : 1912   
## Nurse, Licensed Practical: 1005   
## Pharmacy Technician : 901   
## (Other) : 2481   
## NA's : 4382   
## Action-taken Action-taken-detail  
## Informed staff who made the initial error : 9070 Length:30608   
## Informed staff who was also involved in error: 3703 Class :character   
## Education/Training provided : 2753 Mode :character   
## None : 1921   
## Communication process enhanced : 1670   
## (Other) : 1309   
## NA's :10182   
## Date-record-was-entered  
## Length:30608   
## Class :character   
## Mode :character   
##   
##   
##   
##

count\_day\_week <- newICU1 %>%  
 count(`Day-of-week`)  
  
ggplot(count\_day\_week, aes(x = `Day-of-week`, y = n)) +  
 geom\_bar(stat = 'identity') + geom\_text(aes(label = n, vjust = -.45) )



count\_error\_category <- newICU1 %>%  
 count(`Error-category`)

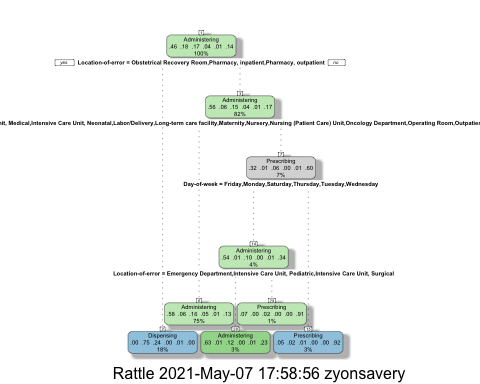
ggplot(count\_error\_category, aes(x = `Error-category`, y = n)) +  
 geom\_bar(stat = 'identity') + geom\_text(aes(label = n, vjust = -.45) )



Improper\_doseICU1 <- subset(newICU1, `Type-of-error` =='Improper dose/quantity')  
Improper\_doseICU1 <- subset(Improper\_doseICU1, select= c(`Type-of-error`, `Error-category`, `Day-of-week`, `Contributing-factor`, `Medication-process-node`, `Location-of-error`))

set.seed(1000)  
Improper\_doseICU1$`Day-of-week` <- as.factor(Improper\_doseICU1$`Day-of-week`)

fit <- rpart( `Medication-process-node` ~ `Day-of-week` + `Location-of-error`, data=Improper\_doseICU1, method="class")   
  
fancyRpartPlot(fit)



pacman::p\_load(tm, wordcloud, e1071, gmodels)

ICU\_data <- subset(newICU2, select= c(`Type-of-error`, `Description-of-error`))  
ICU\_data$isImproperDose <- ICU\_data$`Type-of-error`  
ICU\_data$`Type-of-error` <- factor(ICU\_data$`Type-of-error`)

ICU\_data <- mutate(ICU\_data, isImproperDose = ifelse(as.character(isImproperDose) != "Improper dose/quantity", "Other", as.character(isImproperDose)))  
  
ICU\_data <- subset (ICU\_data, select = -`Type-of-error`)

### Factor

ICU\_data$isImproperDose <- factor(ICU\_data$isImproperDose)

### Frequency

table(ICU\_data$isImproperDose)

##   
## Improper dose/quantity Other   
## 6319 24289

prop.table(table(ICU\_data$isImproperDose))

##   
## Improper dose/quantity Other   
## 0.2064493 0.7935507

## Corpus

corpus <- Corpus(VectorSource(ICU\_data$`Description-of-error`))  
# str(corpus)

corpus[1:10] %>%  
 inspect()

## <<SimpleCorpus>>  
## Metadata: corpus specific: 1, document level (indexed): 0  
## Content: documents: 10  
##   
## [1] Allergy missing on doctor's order   
## [2] Allergy missing on doctor's order   
## [3] Allegy missing on doctor's orders   
## [4] Allegy missing on doctor's orders   
## [5] Pt received celebrex: zantac - ordered on different patient   
## [6] Pt received celebrex: zantac - ordered on different patient   
## [7] Medication arrived late - TPN   
## [8] omission error- MD wrote an order for Lasix 40 mg IV x 1: heplock IV . Order was not pulled or signed off in chart and missed. Noticed at 0200: MD notified and order received to hold Lasix.  
## [9] omission error- MD wrote an order for Lasix 40 mg IV x 1: heplock IV . Order was not pulled or signed off in chart and missed. Noticed at 0200: MD notified and order received to hold Lasix.  
## [10] heparin 25:000 unit bag pulled in error: not hung

### Clean

#### Set to Lower Case

clean\_corpus <- tm\_map(corpus, tolower)

## Warning in tm\_map.SimpleCorpus(corpus, tolower): transformation drops documents

clean\_corpus[1:10] %>%  
 inspect()

## <<SimpleCorpus>>  
## Metadata: corpus specific: 1, document level (indexed): 0  
## Content: documents: 10  
##   
## [1] allergy missing on doctor's order   
## [2] allergy missing on doctor's order   
## [3] allegy missing on doctor's orders   
## [4] allegy missing on doctor's orders   
## [5] pt received celebrex: zantac - ordered on different patient   
## [6] pt received celebrex: zantac - ordered on different patient   
## [7] medication arrived late - tpn   
## [8] omission error- md wrote an order for lasix 40 mg iv x 1: heplock iv . order was not pulled or signed off in chart and missed. noticed at 0200: md notified and order received to hold lasix.  
## [9] omission error- md wrote an order for lasix 40 mg iv x 1: heplock iv . order was not pulled or signed off in chart and missed. noticed at 0200: md notified and order received to hold lasix.  
## [10] heparin 25:000 unit bag pulled in error: not hung

#### Remove Numbers

clean\_corpus1 <- clean\_corpus %>%  
 tm\_map(removeNumbers)

## Warning in tm\_map.SimpleCorpus(., removeNumbers): transformation drops documents

clean\_corpus1[1:10] %>%  
 inspect()

## <<SimpleCorpus>>  
## Metadata: corpus specific: 1, document level (indexed): 0  
## Content: documents: 10  
##   
## [1] allergy missing on doctor's order   
## [2] allergy missing on doctor's order   
## [3] allegy missing on doctor's orders   
## [4] allegy missing on doctor's orders   
## [5] pt received celebrex: zantac - ordered on different patient   
## [6] pt received celebrex: zantac - ordered on different patient   
## [7] medication arrived late - tpn   
## [8] omission error- md wrote an order for lasix mg iv x : heplock iv . order was not pulled or signed off in chart and missed. noticed at : md notified and order received to hold lasix.  
## [9] omission error- md wrote an order for lasix mg iv x : heplock iv . order was not pulled or signed off in chart and missed. noticed at : md notified and order received to hold lasix.  
## [10] heparin : unit bag pulled in error: not hung

#### Remove Stop Words

clean\_corpus2 <- tm\_map(clean\_corpus1, removeWords, stopwords())

## Warning in tm\_map.SimpleCorpus(clean\_corpus1, removeWords, stopwords()):  
## transformation drops documents

clean\_corpus2[1:10] %>%  
 inspect()

## <<SimpleCorpus>>  
## Metadata: corpus specific: 1, document level (indexed): 0  
## Content: documents: 10  
##   
## [1] allergy missing doctor's order   
## [2] allergy missing doctor's order   
## [3] allegy missing doctor's orders   
## [4] allegy missing doctor's orders   
## [5] pt received celebrex: zantac - ordered different patient   
## [6] pt received celebrex: zantac - ordered different patient   
## [7] medication arrived late - tpn   
## [8] omission error- md wrote order lasix mg iv x : heplock iv . order pulled signed chart missed. noticed : md notified order received hold lasix.  
## [9] omission error- md wrote order lasix mg iv x : heplock iv . order pulled signed chart missed. noticed : md notified order received hold lasix.  
## [10] heparin : unit bag pulled error: hung

#### Remove Punctuation

clean\_corpus3 <- tm\_map(clean\_corpus2, removePunctuation)

## Warning in tm\_map.SimpleCorpus(clean\_corpus2, removePunctuation): transformation  
## drops documents

clean\_corpus3[1:10] %>%  
 inspect()

## <<SimpleCorpus>>  
## Metadata: corpus specific: 1, document level (indexed): 0  
## Content: documents: 10  
##   
## [1] allergy missing doctors order   
## [2] allergy missing doctors order   
## [3] allegy missing doctors orders   
## [4] allegy missing doctors orders   
## [5] pt received celebrex zantac ordered different patient   
## [6] pt received celebrex zantac ordered different patient   
## [7] medication arrived late tpn   
## [8] omission error md wrote order lasix mg iv x heplock iv order pulled signed chart missed noticed md notified order received hold lasix  
## [9] omission error md wrote order lasix mg iv x heplock iv order pulled signed chart missed noticed md notified order received hold lasix  
## [10] heparin unit bag pulled error hung

#### Remove White Space

clean\_corpus4 <- tm\_map(clean\_corpus3, stripWhitespace)

## Warning in tm\_map.SimpleCorpus(clean\_corpus3, stripWhitespace): transformation  
## drops documents

clean\_corpus4[1:10] %>%  
 inspect()

## <<SimpleCorpus>>  
## Metadata: corpus specific: 1, document level (indexed): 0  
## Content: documents: 10  
##   
## [1] allergy missing doctors order   
## [2] allergy missing doctors order   
## [3] allegy missing doctors orders   
## [4] allegy missing doctors orders   
## [5] pt received celebrex zantac ordered different patient   
## [6] pt received celebrex zantac ordered different patient   
## [7] medication arrived late tpn   
## [8] omission error md wrote order lasix mg iv x heplock iv order pulled signed chart missed noticed md notified order received hold lasix  
## [9] omission error md wrote order lasix mg iv x heplock iv order pulled signed chart missed noticed md notified order received hold lasix  
## [10] heparin unit bag pulled error hung

### Document Term Matrix from tm package

dmt <- DocumentTermMatrix(clean\_corpus4)  
dmt

## <<DocumentTermMatrix (documents: 32992, terms: 6039)>>  
## Non-/sparse entries: 232268/199006420  
## Sparsity : 100%  
## Maximal term length: 32  
## Weighting : term frequency (tf)

# View(dmt)

## Creating Training and Test Datasets

### Splitting the Data Frame

ICU\_data.max\_size <- nrow(ICU\_data)  
ICU\_data$isImproperDose <- factor(ICU\_data$isImproperDose)  
test.size <- round(ICU\_data.max\_size \* .3)  
train.size <- round(ICU\_data.max\_size \* .7)  
train.df <- ICU\_data[1:train.size,]  
test.df <- ICU\_data[(train.size+1):ICU\_data.max\_size,]

clean\_corpus.max\_size <- length(clean\_corpus4)  
clean\_corpus\_test.size <- round(clean\_corpus.max\_size \* .3)  
clean\_corpus\_train.size <- round(clean\_corpus.max\_size \* .7)  
clean\_corpus\_train.df <- clean\_corpus4[1:clean\_corpus\_train.size ]  
clean\_corpus\_test.df <- clean\_corpus4[(clean\_corpus\_train.size +1):clean\_corpus.max\_size]

### Splitting the Document Term Matrix

dmt.max\_size <- length(dmt)  
dmt\_test.size <- round(dmt.max\_size \* .3)  
dmt\_train.size <- round(dmt.max\_size \* .7)  
dmt\_train <- dmt[1:dmt\_train.size,]  
dmt\_test <- dmt[(dmt\_train.size+1):dmt.max\_size,]

### Compare Frequencies

prop.table(table(ICU\_data$isImproperDose))

##   
## Improper dose/quantity Other   
## 0.2064493 0.7935507

prop.table(table(train.df$isImproperDose))

##   
## Improper dose/quantity Other   
## 0.195537 0.804463

prop.table(table(test.df$isImproperDose))

##   
## Improper dose/quantity Other   
## 0.2315404 0.7684596

### Subset

#### Subset Improper dose/quantity

Improper <- subset(train.df, isImproperDose=="Improper dose/quantity")

#### Subset Other

Other <- subset(train.df, isImproperDose=="Other")

## Find Frequent Terms

### Words That Have a Frequency of 5 or More

Dict <- findFreqTerms(dmt\_train, 1)  
Dict

## [1] "allergy" "doctors" "missing" "order" "allegy" "orders"

matrix\_train <- DocumentTermMatrix(clean\_corpus\_train.df, list(dictionary=Dict))

matrix\_test <- DocumentTermMatrix(clean\_corpus\_test.df, list(dictionary=Dict))

convert\_counts <- function(x) {  
 x <- ifelse(x > 0, 1, 0)   
 x <- factor(x, levels = c(0,1),  
 label= c("Yes","No"))  
 return(x)  
}

### Apply convert\_counts

convert\_train1 <- apply(matrix\_train, MARGIN = 2, convert\_counts)  
convert\_test1 <-apply(matrix\_test, MARGIN = 2, convert\_counts)

#### Check data using str

str(convert\_train1)

## chr [1:23094, 1:6] "No" "No" "Yes" "Yes" "Yes" "Yes" "Yes" "Yes" "Yes" ...  
## - attr(\*, "dimnames")=List of 2  
## ..$ Docs : chr [1:23094] "1" "2" "3" "4" ...  
## ..$ Terms: chr [1:6] "allergy" "doctors" "missing" "order" ...

str(convert\_test1)

## chr [1:9898, 1:6] "Yes" "Yes" "Yes" "Yes" "Yes" "Yes" "Yes" "Yes" "Yes" ...  
## - attr(\*, "dimnames")=List of 2  
## ..$ Docs : chr [1:9898] "1" "2" "3" "4" ...  
## ..$ Terms: chr [1:6] "order" "orders" "doctors" "missing" ...

## Building the Classifier

classifier <- naiveBayes(convert\_train1, train.df$isImproperDose)

## Making Predictions

test\_predict <- predict(classifier, convert\_test1)

## Crosstables

cross\_table <- CrossTable(test\_predict,   
 test.df$isImproperDose,  
 prop.chisq = TRUE,   
 prop.t = TRUE,   
 prop.c = TRUE,   
 prop.r = TRUE,   
 dnn = c('predicted', 'actual'))

##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | N / Table Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 9277   
##   
##   
## | test.df$isImproperDose   
## test\_predict | Improper dose/quantity | Other | Row Total |   
## -------------|------------------------|------------------------|------------------------|  
## Other | 2148 | 7129 | 9277 |   
## | 0.232 | 0.768 | |   
## -------------|------------------------|------------------------|------------------------|  
## Column Total | 2148 | 7129 | 9277 |   
## -------------|------------------------|------------------------|------------------------|  
##   
##