

## 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 x 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" "O
## $ Cause-of-error     : chr [1:30608] "Contraindicated, drug allergy" "Documentation" "Cont
## $ Contributing-factor : chr [1:30608] NA NA NA NA ...
## $ Location-of-error  : chr [1:30608] "Nursing (Patient Care) Unit" "Nursing (Patient Care)
## $ Description-of-error : chr [1:30608] "Allergy missing on doctor's order" "Allergy missing o
## $ 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)

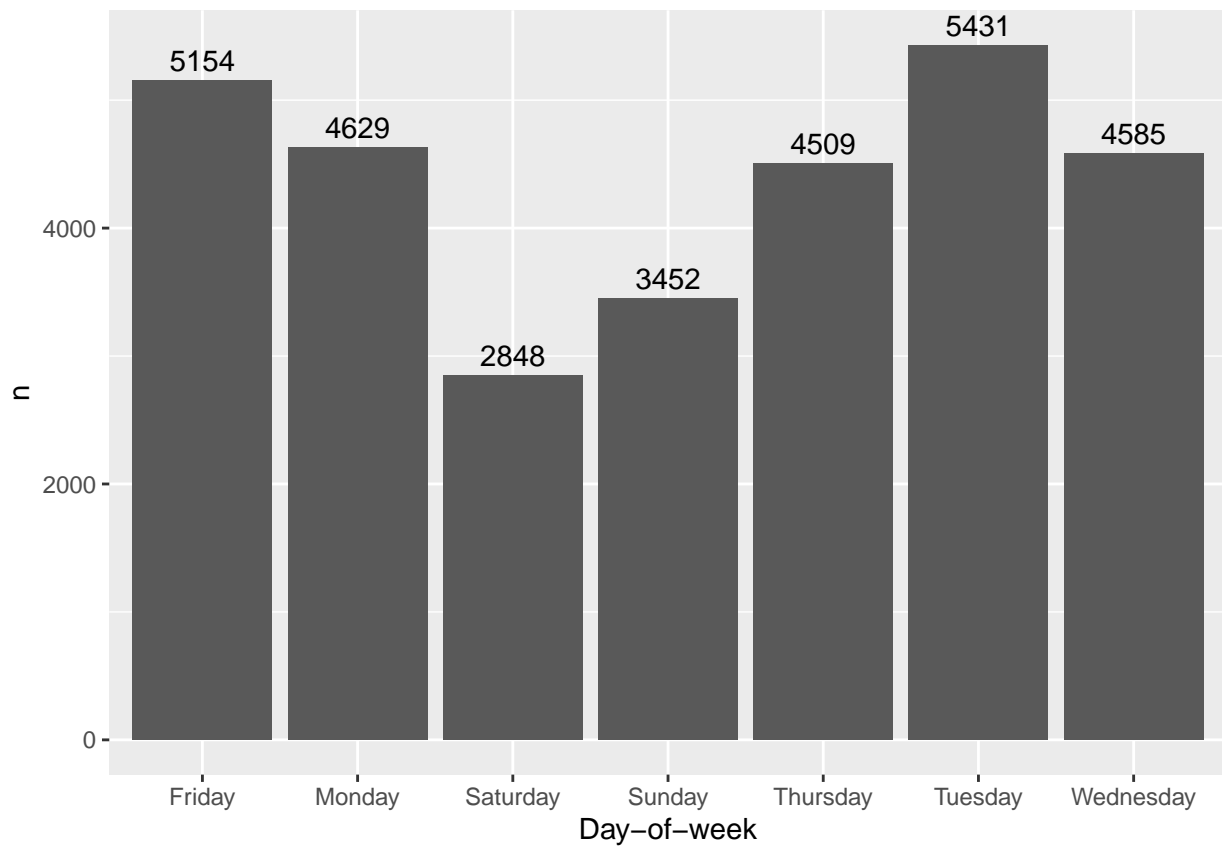
```

##	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
##
##
##
##
```

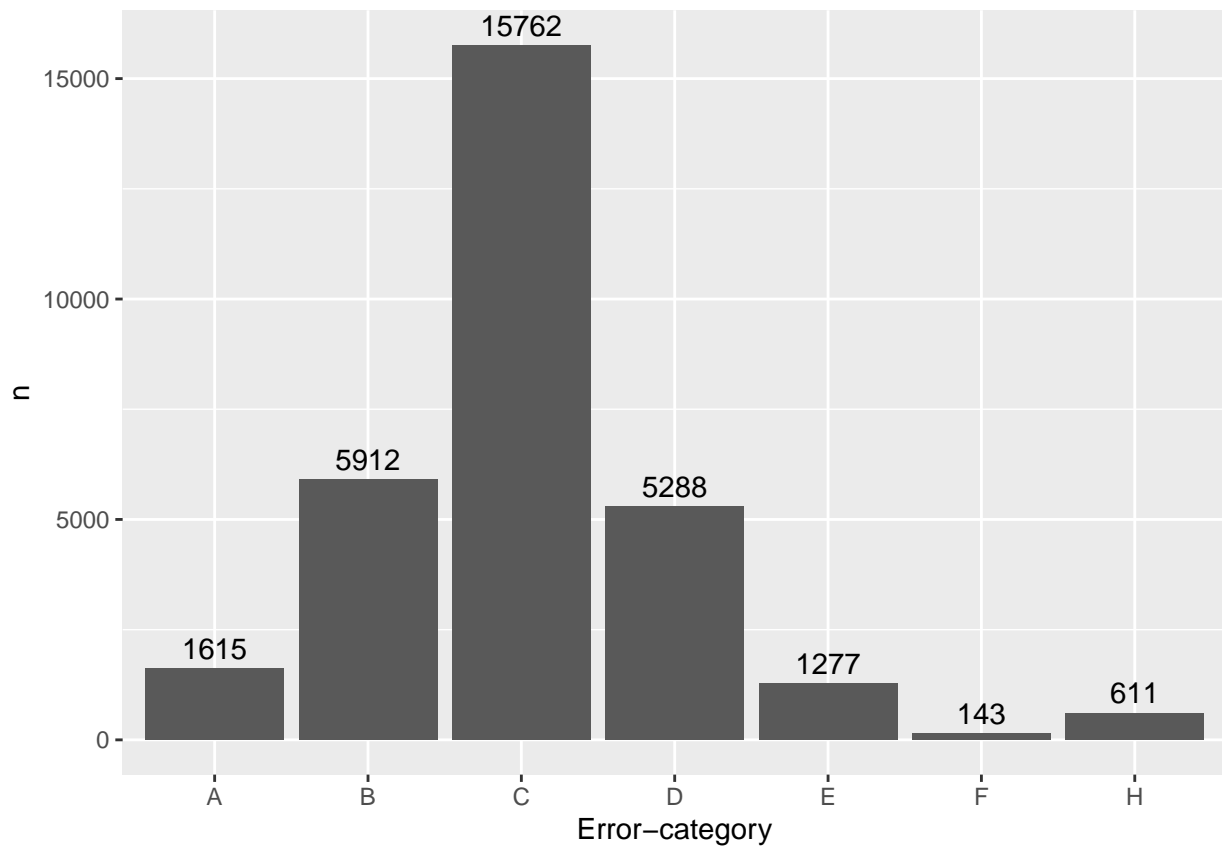
```
count_day_week <- newICU1 %>%
  count(`Day-of-week`)

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`)
```

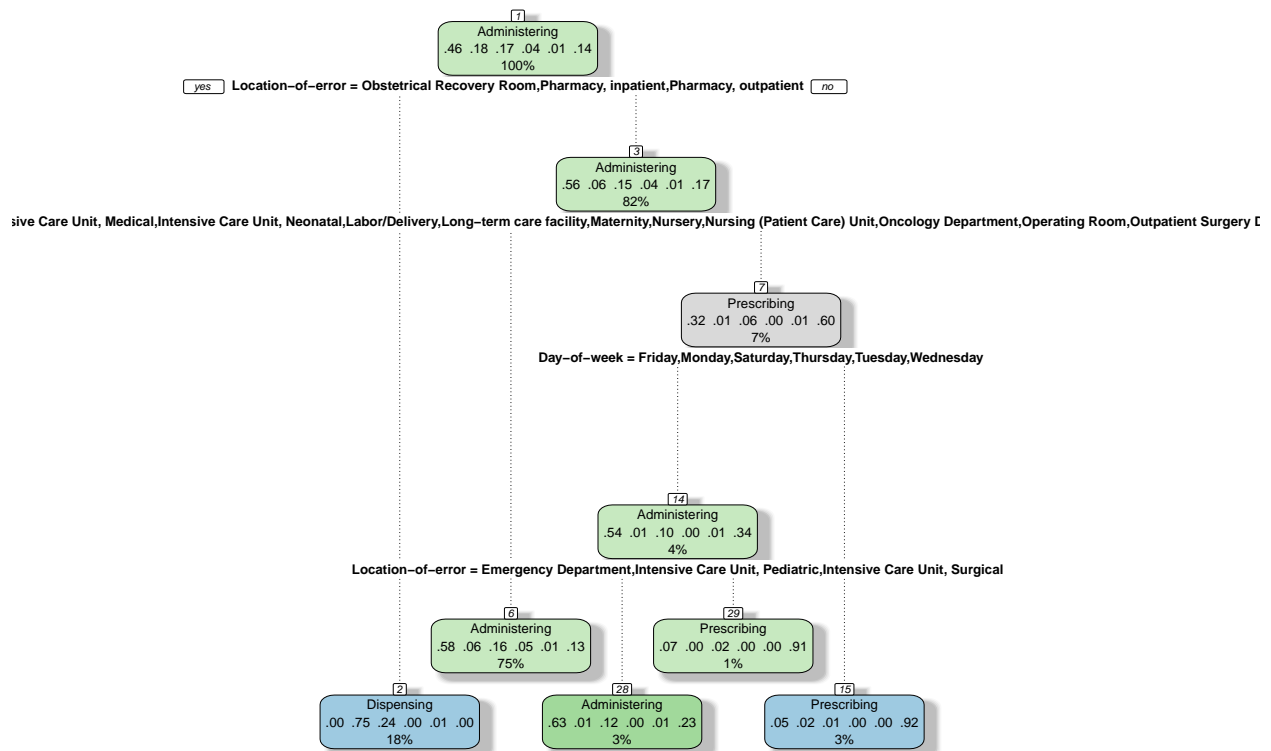
```
ggplot(count_error_category, aes(x = `Error-category`, y = n)) +  
  geom_bar(stat = 'identity') + geom_text(aes(label = n, vjust = -.45) )
```



```
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`))

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,
fancyRpartPlot(fit)
```



Rattle 2021-May-07 18:00:34 zyonsavary

```
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", 1, 0))
```

```
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 s  
## [9] omission error- MD wrote an order for Lasix 40 mg IV x 1: heplock IV . Order was not pulled or s  
## [10] heparin 25:000 unit bag pulled in error: not hung
```

## Clean

```
clean_corpus <- tm_map(corpus, tolower)
```

## Set to Lower Case

```
## 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 s  
## [9] omission error- md wrote an order for lasix 40 mg iv x 1: heplock iv . order was not pulled or s  
## [10] heparin 25:000 unit bag pulled in error: not hung
```

```
clean_corpus1 <- clean_corpus %>%  
  tm_map(removeNumbers)
```

## Remove Numbers

```
## 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 sign
## [9] omission error- md wrote an order for lasix mg iv x : heplock iv . order was not pulled or sign
## [10] heparin : unit bag pulled in error: not hung
```

```
clean_corpus2 <- tm_map(clean_corpus1, removeWords, stopwords())
```

### Remove Stop Words

```
## 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 r
## [9] omission error- md wrote order lasix mg iv x : heplock iv . order pulled signed chart r
## [10] heparin : unit bag pulled error: hung
```

```
clean_corpus3 <- tm_map(clean_corpus2, removePunctuation)
```

### Remove Punctuation

```
## 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 miss
## [9] omission error md wrote order lasix mg iv x heplock iv order pulled signed chart miss
## [10] heparin unit bag pulled error hung
```

```
clean_corpus4 <- tm_map(clean_corpus3, stripWhitespace)
```

### Remove White Space

```
## 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
## [9] omission error md wrote order lasix mg iv x heplock iv order pulled signed chart missed noticed
## [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

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

### Subset Improper dose/quantity

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

### Subset Other

## Find Frequent Terms

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

```
Dict <- findFreqTerms(dmt_train, 1)
Dict

## [1] "allergy" "doctors" "missing" "order" "allegry" "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)
```

```
str(convert_train1)
```

### Check data using str

```
## 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 |
```

```
## -----|-----|-----|-----|
```

```
##
```

```
##
```