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, RColorBrewer, lubridate, dplyr)

#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"

newICU <- ICUEvents  
newICU <- newICU[complete.cases(newICU["Day-of-week"]), ]

* Using R, perform a complete analysis of the dataset (with graphs) representing, for example,
* frequencies of medical errors
* medical errors correlation with
  + day of the week
  + time of the day
  + staff member who made the error
* Since the dependent attribute Type\_of\_Error has multiple values (levels), select one type of error, you will train your classifier
* Check that the error is sufficiently represented in the dataset.
* Clean and prepare your data, using the methods seen in class
* Train your classifier.
* Write a report documenting your work: Dataset description, machine learning method and why it is suitable, Training and testing datasets, and the model evaluation.

# Lets you view document  
# view(newICU)

# Remove useless column  
newICU <- subset(newICU, select= -...11)

#Summarizes doc  
summary(newICU)

## Report-number Facility-ID Error-category Date-of-error   
## Min. : 9475 Min. :11466296 Length:32992 Length:32992   
## 1st Qu.: 14161 1st Qu.:35414425 Class :character Class :character   
## Median : 16865 Median :54716378 Mode :character Mode :character   
## Mean : 20006 Mean :55350119   
## 3rd Qu.: 20340 3rd Qu.:74455222   
## Max. :249347 Max. :99276649   
## Time-of-error Day-of-week Type-of-error Cause-of-error   
## Length:32992 Length:32992 Length:32992 Length:32992   
## 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:32992 Length:32992 Length:32992   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character   
##   
##   
##   
## Medication-process-node Staff-type-initiated-error  
## Length:32992 Length:32992   
## Class :character Class :character   
## Mode :character Mode :character   
##   
##   
##   
## Staff-type-perpetuated-error Staff-type-discovered-error Action-taken   
## Length:32992 Length:32992 Length:32992   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character   
##   
##   
##   
## Action-taken-detail Date-record-was-entered  
## Length:32992 Length:32992   
## Class :character Class :character   
## Mode :character Mode :character   
##   
##   
##

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

## tibble[,18] [32,992 × 18] (S3: tbl\_df/tbl/data.frame)  
## $ Report-number : num [1:32992] 9475 9475 9477 9477 9498 ...  
## $ Facility-ID : num [1:32992] 71778745 71778745 71778745 71778745 33299661 ...  
## $ Error-category : chr [1:32992] "C" "C" "B" "B" ...  
## $ Date-of-error : chr [1:32992] "01/03/2000" "01/03/2000" "01/03/2000" "01/03/2000" ...  
## $ Time-of-error : chr [1:32992] "00:00" "00:00" "14:00" "14:00" ...  
## $ Day-of-week : chr [1:32992] "Monday" "Monday" "Monday" "Monday" ...  
## $ Type-of-error : chr [1:32992] "Omission error" "Omission error" "Omission error" "Omission error" ...  
## $ Cause-of-error : chr [1:32992] "Contraindicated, drug allergy" "Documentation" "Contraindicated, drug allergy" "Documentation" ...  
## $ Contributing-factor : chr [1:32992] NA NA NA NA ...  
## $ Location-of-error : chr [1:32992] "Nursing (Patient Care) Unit" "Nursing (Patient Care) Unit" "Nursing (Patient Care) Unit" "Nursing (Patient Care) Unit" ...  
## $ Description-of-error : chr [1:32992] "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 [1:32992] "Prescribing" "Prescribing" "Documenting" "Documenting" ...  
## $ Staff-type-initiated-error : chr [1:32992] "Physician" "Physician" "Physician" "Physician" ...  
## $ Staff-type-perpetuated-error: chr [1:32992] NA NA NA NA ...  
## $ Staff-type-discovered-error : chr [1:32992] NA NA NA NA ...  
## $ Action-taken : chr [1:32992] NA NA NA NA ...  
## $ Action-taken-detail : chr [1:32992] NA NA NA NA ...  
## $ Date-record-was-entered : chr [1:32992] "01/04/2000" "01/04/2000" "01/04/2000" "01/04/2000" ...

#Displays doc  
newICU

## # A tibble: 32,992 x 18  
## `Report-number` `Facility-ID` `Error-category` `Date-of-error`  
## <dbl> <dbl> <chr> <chr>   
## 1 9475 71778745 C 01/03/2000   
## 2 9475 71778745 C 01/03/2000   
## 3 9477 71778745 B 01/03/2000   
## 4 9477 71778745 B 01/03/2000   
## 5 9498 33299661 C 01/02/2000   
## 6 9498 33299661 C 01/02/2000   
## 7 9509 71778745 C 01/01/2000   
## 8 9542 37744453 C 01/02/2000   
## 9 9542 37744453 C 01/02/2000   
## 10 9618 33299661 B 01/06/2000   
## # … with 32,982 more rows, and 14 more variables: Time-of-error <chr>,  
## # Day-of-week <chr>, Type-of-error <chr>, Cause-of-error <chr>,  
## # Contributing-factor <chr>, Location-of-error <chr>,  
## # Description-of-error <chr>, Medication-process-node <chr>,  
## # Staff-type-initiated-error <chr>, Staff-type-perpetuated-error <chr>,  
## # Staff-type-discovered-error <chr>, Action-taken <chr>,  
## # Action-taken-detail <chr>, Date-record-was-entered <chr>

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

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

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

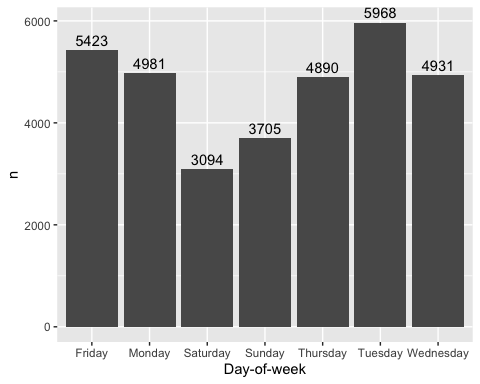
newICU$`Contributing-factor` <- factor(newICU$`Contributing-factor`)  
newICU$`Medication-process-node` <- factor(newICU$`Medication-process-node`)  
newICU$`Staff-type-initiated-error` <- factor(newICU$`Staff-type-initiated-error`)  
newICU$`Staff-type-perpetuated-error` <- factor(newICU$`Staff-type-perpetuated-error`)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

summary(newICU)

## Report-number Facility-ID Error-category Date-of-error   
## Min. : 9475 Min. :11466296 A: 2218 Min. :2020-01-01   
## 1st Qu.: 14161 1st Qu.:35414425 B: 6536 1st Qu.:2020-03-01   
## Median : 16865 Median :54716378 C:16458 Median :2020-06-02   
## Mean : 20006 Mean :55350119 D: 5660 Mean :2020-05-31   
## 3rd Qu.: 20340 3rd Qu.:74455222 E: 1353 3rd Qu.:2020-09-02   
## Max. :249347 Max. :99276649 F: 147 Max. :2020-12-03   
## H: 620 NA's :19079   
## Time-of-error Day-of-week Type-of-error   
## Length:32992 Friday :5423 Omission error :7404   
## Class :character Monday :4981 Improper dose/quantity:6319   
## Mode :character Saturday :3094 Unauthorized drug :3752   
## Sunday :3705 Prescribing error :2255   
## Thursday :4890 Extra dose :2014   
## Tuesday :5968 (Other) :8864   
## Wednesday:4931 NA's :2384   
## Cause-of-error Contributing-factor  
## Performance (human) deficit : 7661 Distractions : 5383   
## Procedure/protocol not followed : 3927 Workload increase : 1912   
## Transcription inaccurate/omitted: 1875 Staff, inexperienced : 1604   
## Communication : 1853 Staffing, insufficient: 1005   
## Knowledge deficit : 1801 Shift change : 971   
## (Other) :13485 (Other) : 5661   
## NA's : 2390 NA's :16456   
## Location-of-error Description-of-error Medication-process-node  
## Length:32992 Length:32992 Administering :12349   
## Class :character Class :character Dispensing : 5976   
## Mode :character Mode :character Documenting : 7614   
## Does not apply: 2218   
## Monitoring : 447   
## Prescribing : 4388   
##   
## Staff-type-initiated-error Staff-type-perpetuated-error  
## Nurse, Registered :13053 Nurse, Registered : 8297   
## Pharmacist : 4672 Pharmacist : 2661   
## Physician : 3048 Pharmacy Technician : 841   
## Pharmacy Technician : 1727 Physician : 805   
## Unit Secretary/Clerk: 1510 Unit Secretary/Clerk: 743   
## (Other) : 6057 (Other) : 2794   
## NA's : 2925 NA's :16851   
## Staff-type-discovered-error  
## Nurse, Registered :16789   
## Pharmacist : 4686   
## Physician : 1968   
## Nurse, Licensed Practical: 1048   
## Pharmacy Technician : 992   
## (Other) : 2659   
## NA's : 4850   
## Action-taken Action-taken-detail  
## Informed staff who made the initial error : 9766 Length:32992   
## Informed staff who was also involved in error: 3975 Class :character   
## Education/Training provided : 2937 Mode :character   
## None : 2119   
## Communication process enhanced : 1812   
## (Other) : 1371   
## NA's :11012   
## Date-record-was-entered  
## Length:32992   
## Class :character   
## Mode :character   
##   
##   
##   
##

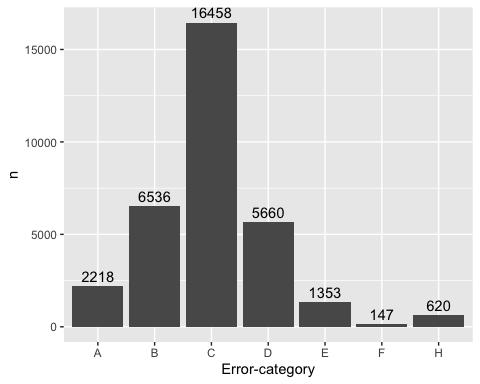
#newICU$`Cause-of-error`

count\_day\_week <- newICU %>%  
 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) )

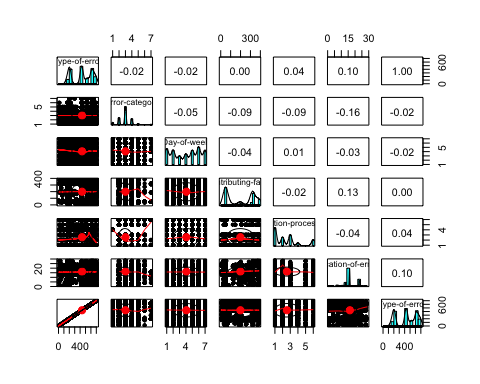


count\_error\_category <- newICU %>%  
 count(`Error-category`)

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

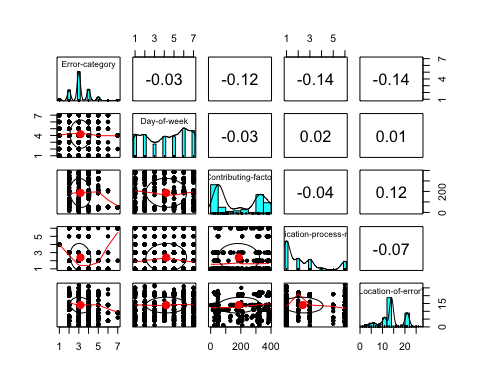


pairs.panels(newICU[c('Type-of-error','Error-category', 'Day-of-week', 'Contributing-factor', 'Medication-process-node', 'Location-of-error', 'Type-of-error')])



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

pairs.panels(Improper\_doseICU[c('Error-category', 'Day-of-week', 'Contributing-factor', 'Medication-process-node', 'Location-of-error')])



set.seed(1000)  
Improper\_doseICU$`Day-of-week` <- as.factor(Improper\_doseICU$`Day-of-week`)  
  
test.size <- round(nrow(Improper\_doseICU) \* .3)  
train.size <- round(nrow(Improper\_doseICU) \* .7)

train.df <- Improper\_doseICU[1:train.size,]  
test.df <- Improper\_doseICU[(train.size+1):(train.size+test.size),]

#### 9. Download the rpart for “Recursive Partitioning and Regression Trees” if it is not in your R base and bring it in your environment.

pacman::p\_load(rpart)

#fit <- rpart(`Medication-process-node` `Day-of-week` + `Error-category`, data=train.df, method="class")   
fit <- rpart( `Location-of-error` ~ `Medication-process-node` + `Day-of-week`, data=train.df, method="class")   
fit

## n=3821 (602 observations deleted due to missingness)  
##   
## node), split, n, loss, yval, (yprob)  
## \* denotes terminal node  
##   
## 1) root 3821 2203 Nursing (Patient Care) Unit (0.0058 0.01 0.036 0.011 0.042 0.016 0.006 0.04 0.0084 0.17 0.0021 0.0031 0.42 0.00026 0.0068 0.0029 0.001 0.001 0.031 0.16 0.024 0.0031 0.0039 0.00079)   
## 2) Medication-process-node=Administering,Documenting,Monitoring,Prescribing 3150 1630 Nursing (Patient Care) Unit (0.007 0.013 0.042 0.013 0.049 0.019 0.0073 0.048 0.0098 0.19 0.0025 0.0038 0.48 0 0.0079 0.0032 0.0013 0.0013 0.037 0.051 0.00095 0.0038 0.0048 0.00095)   
## 4) Day-of-week=Sunday 539 385 Nursing (Patient Care) Unit (0 0 0.074 0 0.12 0.013 0.033 0.24 0.015 0.1 0.0019 0.011 0.29 0 0 0.0019 0 0 0.071 0.039 0 0 0 0)   
## 8) Medication-process-node=Prescribing 246 120 Intensive Care Unit, Surgical (0 0 0.15 0 0.13 0.024 0.073 0.51 0 0 0 0 0.12 0 0 0 0 0 0 0 0 0 0 0) \*  
## 9) Medication-process-node=Administering,Documenting,Monitoring 293 168 Nursing (Patient Care) Unit (0 0 0.014 0 0.11 0.0034 0 0.0068 0.027 0.18 0.0034 0.02 0.43 0 0 0.0034 0 0 0.13 0.072 0 0 0 0) \*  
## 5) Day-of-week=Friday,Monday,Saturday,Thursday,Tuesday,Wednesday 2611 1245 Nursing (Patient Care) Unit (0.0084 0.015 0.035 0.016 0.035 0.02 0.0019 0.0088 0.0088 0.21 0.0027 0.0023 0.52 0 0.0096 0.0034 0.0015 0.0015 0.03 0.054 0.0011 0.0046 0.0057 0.0011)   
## 10) Medication-process-node=Administering 1502 792 Nursing (Patient Care) Unit (0.013 0.002 0.034 0.027 0.04 0.015 0.002 0.015 0.015 0.3 0.004 0.0027 0.47 0 0.0067 0.006 0.002 0.0027 0.022 0 0 0.006 0.0093 0.002)   
## 20) Day-of-week=Friday,Monday,Tuesday 832 455 Long-term care facility (0.013 0.0024 0.0072 0.019 0.035 0.02 0.0036 0.0096 0.018 0.45 0.0048 0.0012 0.34 0 0.012 0 0.0036 0.0048 0.023 0 0 0.011 0.0096 0.0036) \*  
## 21) Day-of-week=Saturday,Thursday,Wednesday 670 247 Nursing (Patient Care) Unit (0.012 0.0015 0.067 0.037 0.046 0.009 0 0.022 0.01 0.11 0.003 0.0045 0.63 0 0 0.013 0 0 0.021 0 0 0 0.009 0) \*  
## 11) Medication-process-node=Documenting,Monitoring,Prescribing 1109 453 Nursing (Patient Care) Unit (0.0027 0.033 0.037 0 0.029 0.027 0.0018 0 0.0009 0.086 0.0009 0.0018 0.59 0 0.014 0 0.0009 0 0.041 0.13 0.0027 0.0027 0.0009 0) \*  
## 3) Medication-process-node=Dispensing 671 234 Pharmacy, inpatient (0 0 0.006 0.0015 0.0075 0 0 0.0015 0.0015 0.046 0 0 0.15 0.0015 0.0015 0.0015 0 0 0 0.65 0.13 0 0 0) \*

pacman::p\_load(rattle, rpart.plot, RColorBrewer)

fancyRpartPlot(fit)

