Exploratory Data Analysis for NSQIP

Vidya Koesmahargyo

11/8/2018

Importing libraries

Importing Dataset (already selected columns)

newdata <- fread("NSQIP.csv", sep="auto")

Dataset structure

nsqip <- newdata  
str(nsqip)

## Classes 'data.table' and 'data.frame': 3974034 obs. of 113 variables:  
## $ V1 : int 1 2 3 4 5 6 7 8 9 10 ...  
## $ X1 : int 482776 482777 482778 482779 482780 482781 482782 482783 482784 482785 ...  
## $ X1\_1 : int 482776 482777 482778 482779 482780 482781 482782 482783 482784 482785 ...  
## $ CaseID : int 966071 966072 966073 966094 966095 966096 966097 966098 966099 966100 ...  
## $ SEX : chr "male" "female" "male" "female" ...  
## $ RACE : logi NA NA NA NA NA NA ...  
## $ PRNCPTX : chr "REPAIR BOWEL OPENING" "MAST, SIMPLE, COMPLETE" "LAPAROSCOPY, APPENDECTOMY" "PARTIAL REMOVAL OF THYROID" ...  
## $ CPT : int 44620 19303 44970 60220 43644 49560 49650 49565 61605 44140 ...  
## $ INOUT : chr "Inpatient" "Inpatient" "Inpatient" "Inpatient" ...  
## $ Age : int 42 77 54 51 50 48 44 55 61 73 ...  
## $ OperYR : int 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 ...  
## $ ANESTHES : chr "General" "General" "General" "General" ...  
## $ SURGSPEC : chr "General Surgery" "General Surgery" "General Surgery" "Otolaryngology (ENT)" ...  
## $ HEIGHT : int 74 64 70 60 66 71 62 62 66 70 ...  
## $ WEIGHT : int 220 160 180 220 282 192 157 231 150 249 ...  
## $ DIABETES : chr "NO" "NO" "NO" "NO" ...  
## $ SMOKE : chr "No" "No" "No" "Yes" ...  
## $ PACKS : int 0 -99 0 20 30 0 0 0 0 0 ...  
## $ ETOH : chr "No" "No" "No" "No" ...  
## $ DYSPNEA : chr "No" "No" "No" "No" ...  
## $ FNSTATUS2 : chr "Independent" "Independent" "Independent" "Independent" ...  
## $ HXCOPD : chr "No" "No" "No" "No" ...  
## $ ASCITES : chr "No" "No" "No" "No" ...  
## $ ESOVAR : chr "No" "No" "No" "No" ...  
## $ HXCHF : chr "No" "No" "No" "No" ...  
## $ HXMI : chr "No" "No" "No" "No" ...  
## $ PRVPCI : chr "No" "No" "No" "No" ...  
## $ PRVPCS : chr "No" "No" "No" "No" ...  
## $ HXANGINA : chr "No" "No" "No" "No" ...  
## $ HYPERMED : chr "Yes" "No" "No" "Yes" ...  
## $ HXPVD : chr "No" "No" "No" "No" ...  
## $ RESTPAIN : chr "No" "No" "No" "No" ...  
## $ RENAFAIL : chr "No" "No" "No" "No" ...  
## $ DIALYSIS : chr "No" "No" "No" "No" ...  
## $ HXTIA : chr "No" "No" "No" "Yes" ...  
## $ CVA : chr "No" "No" "No" "No" ...  
## $ CVANO : chr "No" "No" "No" "No" ...  
## $ DISCANCR : chr "No" "No" "No" "No" ...  
## $ STEROID : chr "No" "No" "No" "No" ...  
## $ WTLOSS : chr "No" "No" "No" "No" ...  
## $ BLEEDDIS : chr "No" "No" "No" "No" ...  
## $ TRANSFUS : chr "No" "No" "No" "No" ...  
## $ CHEMO : chr "No" "No" "No" "No" ...  
## $ PRSEPIS : chr "None" "None" "SIRS" "None" ...  
## $ Pregnancy : chr "No" "No" "No" "No" ...  
## $ PrOper30 : chr "No" "No" "No" "No" ...  
## $ DPRHCT : int 6 -99 0 15 14 -99 7 12 33 6 ...  
## $ DPRPLATE : int 6 -99 0 15 14 -99 7 12 33 6 ...  
## $ PRBUN : num 16 -99 15 11 15 -99 -99 26 18 17 ...  
## $ PRCREAT : num 1.3 0.81 1.12 0.8 0.87 -99 -99 1.05 0.96 1.1 ...  
## $ PRHCT : num 43.6 -99 47.3 35.9 36.1 -99 35.7 48.5 42 41.9 ...  
## $ PRPLATE : num 260 -99 304 364 275 -99 290 212 205 166 ...  
## $ PRPTT : num -99 -99 -99 34.3 23.6 -99 -99 -99 -99 -99 ...  
## $ PRINR : num -99 -99 -99 1.1 1.1 -99 -99 -99 -99 -99 ...  
## $ PRPT : num -99 -99 -99 14.3 14.1 -99 -99 -99 -99 -99 ...  
## $ EMERGNCY : chr "No" "No" "Yes" "No" ...  
## $ ASACLAS : chr "2-Mild Disturb" "2-Mild Disturb" "2-Mild Disturb" "3-Severe Disturb" ...  
## $ RBC : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ OPTIME : int 211 97 65 92 114 62 10 53 330 166 ...  
## $ TYPEINTOC : chr "NULL" "NULL" "NULL" "NULL" ...  
## $ SDISDT : int 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 ...  
## $ HDISDT : int 2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 ...  
## $ YRDEATH : int -99 -99 -99 -99 -99 -99 -99 -99 -99 -99 ...  
## $ TOTHLOS : int 5 1 1 1 2 0 0 0 1 12 ...  
## $ HtoODay : num 0 0 0 0 0 0 0 0 0 0 ...  
## $ TOTSLOS : int 5 1 1 1 2 0 0 0 1 12 ...  
## $ NPULEMBOL : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ PULEMBOL : chr "No Complication" "No Complication" "No Complication" "No Complication" ...  
## $ DPULEMBOL : int -99 -99 -99 -99 -99 -99 -99 -99 -99 -99 ...  
## $ NRENAINSF : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ RENAINSF : chr "No Complication" "No Complication" "No Complication" "No Complication" ...  
## $ DRENAINSF : int -99 -99 -99 -99 -99 -99 -99 -99 -99 -99 ...  
## $ NOPRENAFL : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ OPRENAFL : chr "No Complication" "No Complication" "No Complication" "No Complication" ...  
## $ DOPRENAFL : int -99 -99 -99 -99 -99 -99 -99 -99 -99 -99 ...  
## $ NCNSCVA : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ CNSCVA : chr "No Complication" "No Complication" "No Complication" "No Complication" ...  
## $ DCNSCVA : int -99 -99 -99 -99 -99 -99 -99 -99 -99 -99 ...  
## $ NCDARREST : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ CDARREST : chr "No Complication" "No Complication" "No Complication" "No Complication" ...  
## $ DCDARREST : int -99 -99 -99 -99 -99 -99 -99 -99 -99 -99 ...  
## $ NCDMI : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ CDMI : chr "No Complication" "No Complication" "No Complication" "No Complication" ...  
## $ DCDMI : int -99 -99 -99 -99 -99 -99 -99 -99 -99 -99 ...  
## $ NOTHBLEED : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ OTHBLEED : chr "No Complication" "No Complication" "No Complication" "No Complication" ...  
## $ DOTHBLEED : num -99 -99 -99 -99 -99 -99 -99 -99 -99 -99 ...  
## $ NOTHDVT : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ OTHDVT : chr "No Complication" "No Complication" "No Complication" "No Complication" ...  
## $ DOTHDVT : int -99 -99 -99 -99 -99 -99 -99 -99 -99 -99 ...  
## $ PODIAG : chr "562.03" "174.4" "540.9" "246.2" ...  
## $ RETURNOR : chr "No" "No" "No" "No" ...  
## $ DOpertoD : int -99 -99 -99 -99 -99 -99 -99 -99 -99 -99 ...  
## $ DOptoDis : int 5 1 1 1 2 0 0 0 1 12 ...  
## $ MorbProb\_A : logi NA NA NA NA NA NA ...  
## $ MortProb\_A : logi NA NA NA NA NA NA ...  
## $ RACE\_NEW : chr "White" "White" "White" "Black or African American" ...  
## $ ETHNICITY\_HISPANIC: chr "No" "No" "No" "No" ...  
## $ MORTPROB : num 0.00127 0.00104 0.00131 NA 0.00156 ...  
## [list output truncated]  
## - attr(\*, ".internal.selfref")=<externalptr>

Look at non-cardiac surgeries only.

#select out cardiac surgery  
nsqip <- filter(nsqip, SURGSPEC != "Cardiac Surgery")  
colnames <- colnames(nsqip)

Change thecharacter variables into factors for plotting/analysis purposes

Examine how much of the data set is missing, explicitlylt (NAs) or implicitly (-99).

#find percent NAs  
na\_lst <- sapply(nsqip, function(x) sum(is.na(x)))  
na\_lst

## V1 X1 X1\_1   
## 0 0 881469   
## CaseID SEX RACE   
## 0 0 3950083   
## PRNCPTX CPT INOUT   
## 0 1534 0   
## Age OperYR ANESTHES   
## 45381 0 0   
## SURGSPEC HEIGHT WEIGHT   
## 0 0 0   
## DIABETES SMOKE PACKS   
## 0 0 881469   
## ETOH DYSPNEA FNSTATUS2   
## 881469 0 0   
## HXCOPD ASCITES ESOVAR   
## 0 0 881469   
## HXCHF HXMI PRVPCI   
## 0 881469 881469   
## PRVPCS HXANGINA HYPERMED   
## 881469 881469 0   
## HXPVD RESTPAIN RENAFAIL   
## 881469 881469 0   
## DIALYSIS HXTIA CVA   
## 0 881469 881469   
## CVANO DISCANCR STEROID   
## 881469 0 0   
## WTLOSS BLEEDDIS TRANSFUS   
## 0 0 0   
## CHEMO PRSEPIS Pregnancy   
## 881469 0 881469   
## PrOper30 DPRHCT DPRPLATE   
## 881469 0 0   
## PRBUN PRCREAT PRHCT   
## 0 0 0   
## PRPLATE PRPTT PRINR   
## 0 0 0   
## PRPT EMERGNCY ASACLAS   
## 0 0 0   
## RBC OPTIME TYPEINTOC   
## 881469 0 881469   
## SDISDT HDISDT YRDEATH   
## 881469 0 0   
## TOTHLOS HtoODay TOTSLOS   
## 0 0 881469   
## NPULEMBOL PULEMBOL DPULEMBOL   
## 0 0 0   
## NRENAINSF RENAINSF DRENAINSF   
## 0 0 0   
## NOPRENAFL OPRENAFL DOPRENAFL   
## 0 0 0   
## NCNSCVA CNSCVA DCNSCVA   
## 0 0 0   
## NCDARREST CDARREST DCDARREST   
## 0 0 0   
## NCDMI CDMI DCDMI   
## 0 0 0   
## NOTHBLEED OTHBLEED DOTHBLEED   
## 0 0 0   
## NOTHDVT OTHDVT DOTHDVT   
## 0 0 0   
## PODIAG RETURNOR DOpertoD   
## 0 0 36   
## DOptoDis MorbProb\_A MortProb\_A   
## 0 3950083 3950083   
## RACE\_NEW ETHNICITY\_HISPANIC MORTPROB   
## 0 0 65310   
## MORBPROB ANESURG SURGANE   
## 65310 881469 881469   
## ELECTSURG DISCHDEST READMISSION   
## 693992 693992 1575461   
## UNPLANREADMISSION REOPERATION STILLINHOSP   
## 1575461 1575461 693992   
## REOPERATION1 RETORPODAYS REOPORCPT1   
## 1132666 1132666 1132666   
## RETORRELATED REOPORICD91   
## 1132666 1132666

percent\_na <- lapply(na\_lst, function(x) percent((x/nrow(nsqip))))  
percent\_na

## $V1  
## [1] "0%"  
##   
## $X1  
## [1] "0%"  
##   
## $X1\_1  
## [1] "22.3%"  
##   
## $CaseID  
## [1] "0%"  
##   
## $SEX  
## [1] "0%"  
##   
## $RACE  
## [1] "100%"  
##   
## $PRNCPTX  
## [1] "0%"  
##   
## $CPT  
## [1] "0.0388%"  
##   
## $INOUT  
## [1] "0%"  
##   
## $Age  
## [1] "1.15%"  
##   
## $OperYR  
## [1] "0%"  
##   
## $ANESTHES  
## [1] "0%"  
##   
## $SURGSPEC  
## [1] "0%"  
##   
## $HEIGHT  
## [1] "0%"  
##   
## $WEIGHT  
## [1] "0%"  
##   
## $DIABETES  
## [1] "0%"  
##   
## $SMOKE  
## [1] "0%"  
##   
## $PACKS  
## [1] "22.3%"  
##   
## $ETOH  
## [1] "22.3%"  
##   
## $DYSPNEA  
## [1] "0%"  
##   
## $FNSTATUS2  
## [1] "0%"  
##   
## $HXCOPD  
## [1] "0%"  
##   
## $ASCITES  
## [1] "0%"  
##   
## $ESOVAR  
## [1] "22.3%"  
##   
## $HXCHF  
## [1] "0%"  
##   
## $HXMI  
## [1] "22.3%"  
##   
## $PRVPCI  
## [1] "22.3%"  
##   
## $PRVPCS  
## [1] "22.3%"  
##   
## $HXANGINA  
## [1] "22.3%"  
##   
## $HYPERMED  
## [1] "0%"  
##   
## $HXPVD  
## [1] "22.3%"  
##   
## $RESTPAIN  
## [1] "22.3%"  
##   
## $RENAFAIL  
## [1] "0%"  
##   
## $DIALYSIS  
## [1] "0%"  
##   
## $HXTIA  
## [1] "22.3%"  
##   
## $CVA  
## [1] "22.3%"  
##   
## $CVANO  
## [1] "22.3%"  
##   
## $DISCANCR  
## [1] "0%"  
##   
## $STEROID  
## [1] "0%"  
##   
## $WTLOSS  
## [1] "0%"  
##   
## $BLEEDDIS  
## [1] "0%"  
##   
## $TRANSFUS  
## [1] "0%"  
##   
## $CHEMO  
## [1] "22.3%"  
##   
## $PRSEPIS  
## [1] "0%"  
##   
## $Pregnancy  
## [1] "22.3%"  
##   
## $PrOper30  
## [1] "22.3%"  
##   
## $DPRHCT  
## [1] "0%"  
##   
## $DPRPLATE  
## [1] "0%"  
##   
## $PRBUN  
## [1] "0%"  
##   
## $PRCREAT  
## [1] "0%"  
##   
## $PRHCT  
## [1] "0%"  
##   
## $PRPLATE  
## [1] "0%"  
##   
## $PRPTT  
## [1] "0%"  
##   
## $PRINR  
## [1] "0%"  
##   
## $PRPT  
## [1] "0%"  
##   
## $EMERGNCY  
## [1] "0%"  
##   
## $ASACLAS  
## [1] "0%"  
##   
## $RBC  
## [1] "22.3%"  
##   
## $OPTIME  
## [1] "0%"  
##   
## $TYPEINTOC  
## [1] "22.3%"  
##   
## $SDISDT  
## [1] "22.3%"  
##   
## $HDISDT  
## [1] "0%"  
##   
## $YRDEATH  
## [1] "0%"  
##   
## $TOTHLOS  
## [1] "0%"  
##   
## $HtoODay  
## [1] "0%"  
##   
## $TOTSLOS  
## [1] "22.3%"  
##   
## $NPULEMBOL  
## [1] "0%"  
##   
## $PULEMBOL  
## [1] "0%"  
##   
## $DPULEMBOL  
## [1] "0%"  
##   
## $NRENAINSF  
## [1] "0%"  
##   
## $RENAINSF  
## [1] "0%"  
##   
## $DRENAINSF  
## [1] "0%"  
##   
## $NOPRENAFL  
## [1] "0%"  
##   
## $OPRENAFL  
## [1] "0%"  
##   
## $DOPRENAFL  
## [1] "0%"  
##   
## $NCNSCVA  
## [1] "0%"  
##   
## $CNSCVA  
## [1] "0%"  
##   
## $DCNSCVA  
## [1] "0%"  
##   
## $NCDARREST  
## [1] "0%"  
##   
## $CDARREST  
## [1] "0%"  
##   
## $DCDARREST  
## [1] "0%"  
##   
## $NCDMI  
## [1] "0%"  
##   
## $CDMI  
## [1] "0%"  
##   
## $DCDMI  
## [1] "0%"  
##   
## $NOTHBLEED  
## [1] "0%"  
##   
## $OTHBLEED  
## [1] "0%"  
##   
## $DOTHBLEED  
## [1] "0%"  
##   
## $NOTHDVT  
## [1] "0%"  
##   
## $OTHDVT  
## [1] "0%"  
##   
## $DOTHDVT  
## [1] "0%"  
##   
## $PODIAG  
## [1] "0%"  
##   
## $RETURNOR  
## [1] "0%"  
##   
## $DOpertoD  
## [1] "0.000911%"  
##   
## $DOptoDis  
## [1] "0%"  
##   
## $MorbProb\_A  
## [1] "100%"  
##   
## $MortProb\_A  
## [1] "100%"  
##   
## $RACE\_NEW  
## [1] "0%"  
##   
## $ETHNICITY\_HISPANIC  
## [1] "0%"  
##   
## $MORTPROB  
## [1] "1.65%"  
##   
## $MORBPROB  
## [1] "1.65%"  
##   
## $ANESURG  
## [1] "22.3%"  
##   
## $SURGANE  
## [1] "22.3%"  
##   
## $ELECTSURG  
## [1] "17.6%"  
##   
## $DISCHDEST  
## [1] "17.6%"  
##   
## $READMISSION  
## [1] "39.9%"  
##   
## $UNPLANREADMISSION  
## [1] "39.9%"  
##   
## $REOPERATION  
## [1] "39.9%"  
##   
## $STILLINHOSP  
## [1] "17.6%"  
##   
## $REOPERATION1  
## [1] "28.7%"  
##   
## $RETORPODAYS  
## [1] "28.7%"  
##   
## $REOPORCPT1  
## [1] "28.7%"  
##   
## $RETORRELATED  
## [1] "28.7%"  
##   
## $REOPORICD91  
## [1] "28.7%"

#implicit NAs/ = '-99' list for each column  
implicit\_lst <- sapply(nsqip, function(x) sum(x == -99))  
implicit\_lst

## V1 X1 X1\_1   
## 0 0 NA   
## CaseID SEX RACE   
## 0 0 NA   
## PRNCPTX CPT INOUT   
## 0 NA 0   
## Age OperYR ANESTHES   
## NA 0 0   
## SURGSPEC HEIGHT WEIGHT   
## 0 80100 42418   
## DIABETES SMOKE PACKS   
## 0 0 NA   
## ETOH DYSPNEA FNSTATUS2   
## NA 0 0   
## HXCOPD ASCITES ESOVAR   
## 0 0 NA   
## HXCHF HXMI PRVPCI   
## 0 NA NA   
## PRVPCS HXANGINA HYPERMED   
## NA NA 0   
## HXPVD RESTPAIN RENAFAIL   
## NA NA 0   
## DIALYSIS HXTIA CVA   
## 0 NA NA   
## CVANO DISCANCR STEROID   
## NA 0 0   
## WTLOSS BLEEDDIS TRANSFUS   
## 0 0 0   
## CHEMO PRSEPIS Pregnancy   
## NA 0 NA   
## PrOper30 DPRHCT DPRPLATE   
## NA 551613 620528   
## PRBUN PRCREAT PRHCT   
## 847478 710485 554237   
## PRPLATE PRPTT PRINR   
## 616020 2572622 2246940   
## PRPT EMERGNCY ASACLAS   
## 3488413 0 0   
## RBC OPTIME TYPEINTOC   
## NA 642 NA   
## SDISDT HDISDT YRDEATH   
## NA 2218 3897834   
## TOTHLOS HtoODay TOTSLOS   
## 3251 236 NA   
## NPULEMBOL PULEMBOL DPULEMBOL   
## 0 0 3936984   
## NRENAINSF RENAINSF DRENAINSF   
## 0 0 3939143   
## NOPRENAFL OPRENAFL DOPRENAFL   
## 0 0 3937554   
## NCNSCVA CNSCVA DCNSCVA   
## 0 0 3942202   
## NCDARREST CDARREST DCDARREST   
## 0 0 3937519   
## NCDMI CDMI DCDMI   
## 0 0 3935397   
## NOTHBLEED OTHBLEED DOTHBLEED   
## 0 0 3732556   
## NOTHDVT OTHDVT DOTHDVT   
## 0 0 3926404   
## PODIAG RETURNOR DOpertoD   
## 0 0 NA   
## DOptoDis MorbProb\_A MortProb\_A   
## 2224 NA NA   
## RACE\_NEW ETHNICITY\_HISPANIC MORTPROB   
## 0 1 NA   
## MORBPROB ANESURG SURGANE   
## NA NA NA   
## ELECTSURG DISCHDEST READMISSION   
## NA NA NA   
## UNPLANREADMISSION REOPERATION STILLINHOSP   
## NA NA NA   
## REOPERATION1 RETORPODAYS REOPORCPT1   
## NA NA NA   
## RETORRELATED REOPORICD91   
## NA NA

#percent implicit for each column (made NAs, since they were null values in the beginning)  
implicit\_lst[is.na(implicit\_lst)] <- 0  
percent\_implicit <- sapply(implicit\_lst, function(x) percent(x/nrow(newdata)))  
percent\_implicit

## V1 X1 X1\_1   
## "0%" "0%" "0%"   
## CaseID SEX RACE   
## "0%" "0%" "0%"   
## PRNCPTX CPT INOUT   
## "0%" "0%" "0%"   
## Age OperYR ANESTHES   
## "0%" "0%" "0%"   
## SURGSPEC HEIGHT WEIGHT   
## "0%" "2.02%" "1.07%"   
## DIABETES SMOKE PACKS   
## "0%" "0%" "0%"   
## ETOH DYSPNEA FNSTATUS2   
## "0%" "0%" "0%"   
## HXCOPD ASCITES ESOVAR   
## "0%" "0%" "0%"   
## HXCHF HXMI PRVPCI   
## "0%" "0%" "0%"   
## PRVPCS HXANGINA HYPERMED   
## "0%" "0%" "0%"   
## HXPVD RESTPAIN RENAFAIL   
## "0%" "0%" "0%"   
## DIALYSIS HXTIA CVA   
## "0%" "0%" "0%"   
## CVANO DISCANCR STEROID   
## "0%" "0%" "0%"   
## WTLOSS BLEEDDIS TRANSFUS   
## "0%" "0%" "0%"   
## CHEMO PRSEPIS Pregnancy   
## "0%" "0%" "0%"   
## PrOper30 DPRHCT DPRPLATE   
## "0%" "13.9%" "15.6%"   
## PRBUN PRCREAT PRHCT   
## "21.3%" "17.9%" "13.9%"   
## PRPLATE PRPTT PRINR   
## "15.5%" "64.7%" "56.5%"   
## PRPT EMERGNCY ASACLAS   
## "87.8%" "0%" "0%"   
## RBC OPTIME TYPEINTOC   
## "0%" "0.0162%" "0%"   
## SDISDT HDISDT YRDEATH   
## "0%" "0.0558%" "98.1%"   
## TOTHLOS HtoODay TOTSLOS   
## "0.0818%" "0.00594%" "0%"   
## NPULEMBOL PULEMBOL DPULEMBOL   
## "0%" "0%" "99.1%"   
## NRENAINSF RENAINSF DRENAINSF   
## "0%" "0%" "99.1%"   
## NOPRENAFL OPRENAFL DOPRENAFL   
## "0%" "0%" "99.1%"   
## NCNSCVA CNSCVA DCNSCVA   
## "0%" "0%" "99.2%"   
## NCDARREST CDARREST DCDARREST   
## "0%" "0%" "99.1%"   
## NCDMI CDMI DCDMI   
## "0%" "0%" "99.0%"   
## NOTHBLEED OTHBLEED DOTHBLEED   
## "0%" "0%" "93.9%"   
## NOTHDVT OTHDVT DOTHDVT   
## "0%" "0%" "98.8%"   
## PODIAG RETURNOR DOpertoD   
## "0%" "0%" "0%"   
## DOptoDis MorbProb\_A MortProb\_A   
## "0.0560%" "0%" "0%"   
## RACE\_NEW ETHNICITY\_HISPANIC MORTPROB   
## "0%" "0.0000252%" "0%"   
## MORBPROB ANESURG SURGANE   
## "0%" "0%" "0%"   
## ELECTSURG DISCHDEST READMISSION   
## "0%" "0%" "0%"   
## UNPLANREADMISSION REOPERATION STILLINHOSP   
## "0%" "0%" "0%"   
## REOPERATION1 RETORPODAYS REOPORCPT1   
## "0%" "0%" "0%"   
## RETORRELATED REOPORICD91   
## "0%" "0%"

#find all missing values  
missing\_lst <- implicit\_lst + na\_lst  
missing\_lst

## V1 X1 X1\_1   
## 0 0 881469   
## CaseID SEX RACE   
## 0 0 3950083   
## PRNCPTX CPT INOUT   
## 0 1534 0   
## Age OperYR ANESTHES   
## 45381 0 0   
## SURGSPEC HEIGHT WEIGHT   
## 0 80100 42418   
## DIABETES SMOKE PACKS   
## 0 0 881469   
## ETOH DYSPNEA FNSTATUS2   
## 881469 0 0   
## HXCOPD ASCITES ESOVAR   
## 0 0 881469   
## HXCHF HXMI PRVPCI   
## 0 881469 881469   
## PRVPCS HXANGINA HYPERMED   
## 881469 881469 0   
## HXPVD RESTPAIN RENAFAIL   
## 881469 881469 0   
## DIALYSIS HXTIA CVA   
## 0 881469 881469   
## CVANO DISCANCR STEROID   
## 881469 0 0   
## WTLOSS BLEEDDIS TRANSFUS   
## 0 0 0   
## CHEMO PRSEPIS Pregnancy   
## 881469 0 881469   
## PrOper30 DPRHCT DPRPLATE   
## 881469 551613 620528   
## PRBUN PRCREAT PRHCT   
## 847478 710485 554237   
## PRPLATE PRPTT PRINR   
## 616020 2572622 2246940   
## PRPT EMERGNCY ASACLAS   
## 3488413 0 0   
## RBC OPTIME TYPEINTOC   
## 881469 642 881469   
## SDISDT HDISDT YRDEATH   
## 881469 2218 3897834   
## TOTHLOS HtoODay TOTSLOS   
## 3251 236 881469   
## NPULEMBOL PULEMBOL DPULEMBOL   
## 0 0 3936984   
## NRENAINSF RENAINSF DRENAINSF   
## 0 0 3939143   
## NOPRENAFL OPRENAFL DOPRENAFL   
## 0 0 3937554   
## NCNSCVA CNSCVA DCNSCVA   
## 0 0 3942202   
## NCDARREST CDARREST DCDARREST   
## 0 0 3937519   
## NCDMI CDMI DCDMI   
## 0 0 3935397   
## NOTHBLEED OTHBLEED DOTHBLEED   
## 0 0 3732556   
## NOTHDVT OTHDVT DOTHDVT   
## 0 0 3926404   
## PODIAG RETURNOR DOpertoD   
## 0 0 36   
## DOptoDis MorbProb\_A MortProb\_A   
## 2224 3950083 3950083   
## RACE\_NEW ETHNICITY\_HISPANIC MORTPROB   
## 0 1 65310   
## MORBPROB ANESURG SURGANE   
## 65310 881469 881469   
## ELECTSURG DISCHDEST READMISSION   
## 693992 693992 1575461   
## UNPLANREADMISSION REOPERATION STILLINHOSP   
## 1575461 1575461 693992   
## REOPERATION1 RETORPODAYS REOPORCPT1   
## 1132666 1132666 1132666   
## RETORRELATED REOPORICD91   
## 1132666 1132666

percent\_missing <- lapply(missing\_lst, function(x) percent((x/nrow(nsqip))))  
percent\_missing

## $V1  
## [1] "0%"  
##   
## $X1  
## [1] "0%"  
##   
## $X1\_1  
## [1] "22.3%"  
##   
## $CaseID  
## [1] "0%"  
##   
## $SEX  
## [1] "0%"  
##   
## $RACE  
## [1] "100%"  
##   
## $PRNCPTX  
## [1] "0%"  
##   
## $CPT  
## [1] "0.0388%"  
##   
## $INOUT  
## [1] "0%"  
##   
## $Age  
## [1] "1.15%"  
##   
## $OperYR  
## [1] "0%"  
##   
## $ANESTHES  
## [1] "0%"  
##   
## $SURGSPEC  
## [1] "0%"  
##   
## $HEIGHT  
## [1] "2.03%"  
##   
## $WEIGHT  
## [1] "1.07%"  
##   
## $DIABETES  
## [1] "0%"  
##   
## $SMOKE  
## [1] "0%"  
##   
## $PACKS  
## [1] "22.3%"  
##   
## $ETOH  
## [1] "22.3%"  
##   
## $DYSPNEA  
## [1] "0%"  
##   
## $FNSTATUS2  
## [1] "0%"  
##   
## $HXCOPD  
## [1] "0%"  
##   
## $ASCITES  
## [1] "0%"  
##   
## $ESOVAR  
## [1] "22.3%"  
##   
## $HXCHF  
## [1] "0%"  
##   
## $HXMI  
## [1] "22.3%"  
##   
## $PRVPCI  
## [1] "22.3%"  
##   
## $PRVPCS  
## [1] "22.3%"  
##   
## $HXANGINA  
## [1] "22.3%"  
##   
## $HYPERMED  
## [1] "0%"  
##   
## $HXPVD  
## [1] "22.3%"  
##   
## $RESTPAIN  
## [1] "22.3%"  
##   
## $RENAFAIL  
## [1] "0%"  
##   
## $DIALYSIS  
## [1] "0%"  
##   
## $HXTIA  
## [1] "22.3%"  
##   
## $CVA  
## [1] "22.3%"  
##   
## $CVANO  
## [1] "22.3%"  
##   
## $DISCANCR  
## [1] "0%"  
##   
## $STEROID  
## [1] "0%"  
##   
## $WTLOSS  
## [1] "0%"  
##   
## $BLEEDDIS  
## [1] "0%"  
##   
## $TRANSFUS  
## [1] "0%"  
##   
## $CHEMO  
## [1] "22.3%"  
##   
## $PRSEPIS  
## [1] "0%"  
##   
## $Pregnancy  
## [1] "22.3%"  
##   
## $PrOper30  
## [1] "22.3%"  
##   
## $DPRHCT  
## [1] "14.0%"  
##   
## $DPRPLATE  
## [1] "15.7%"  
##   
## $PRBUN  
## [1] "21.5%"  
##   
## $PRCREAT  
## [1] "18.0%"  
##   
## $PRHCT  
## [1] "14.0%"  
##   
## $PRPLATE  
## [1] "15.6%"  
##   
## $PRPTT  
## [1] "65.1%"  
##   
## $PRINR  
## [1] "56.9%"  
##   
## $PRPT  
## [1] "88.3%"  
##   
## $EMERGNCY  
## [1] "0%"  
##   
## $ASACLAS  
## [1] "0%"  
##   
## $RBC  
## [1] "22.3%"  
##   
## $OPTIME  
## [1] "0.0163%"  
##   
## $TYPEINTOC  
## [1] "22.3%"  
##   
## $SDISDT  
## [1] "22.3%"  
##   
## $HDISDT  
## [1] "0.0562%"  
##   
## $YRDEATH  
## [1] "98.7%"  
##   
## $TOTHLOS  
## [1] "0.0823%"  
##   
## $HtoODay  
## [1] "0.00597%"  
##   
## $TOTSLOS  
## [1] "22.3%"  
##   
## $NPULEMBOL  
## [1] "0%"  
##   
## $PULEMBOL  
## [1] "0%"  
##   
## $DPULEMBOL  
## [1] "99.7%"  
##   
## $NRENAINSF  
## [1] "0%"  
##   
## $RENAINSF  
## [1] "0%"  
##   
## $DRENAINSF  
## [1] "99.7%"  
##   
## $NOPRENAFL  
## [1] "0%"  
##   
## $OPRENAFL  
## [1] "0%"  
##   
## $DOPRENAFL  
## [1] "99.7%"  
##   
## $NCNSCVA  
## [1] "0%"  
##   
## $CNSCVA  
## [1] "0%"  
##   
## $DCNSCVA  
## [1] "99.8%"  
##   
## $NCDARREST  
## [1] "0%"  
##   
## $CDARREST  
## [1] "0%"  
##   
## $DCDARREST  
## [1] "99.7%"  
##   
## $NCDMI  
## [1] "0%"  
##   
## $CDMI  
## [1] "0%"  
##   
## $DCDMI  
## [1] "99.6%"  
##   
## $NOTHBLEED  
## [1] "0%"  
##   
## $OTHBLEED  
## [1] "0%"  
##   
## $DOTHBLEED  
## [1] "94.5%"  
##   
## $NOTHDVT  
## [1] "0%"  
##   
## $OTHDVT  
## [1] "0%"  
##   
## $DOTHDVT  
## [1] "99.4%"  
##   
## $PODIAG  
## [1] "0%"  
##   
## $RETURNOR  
## [1] "0%"  
##   
## $DOpertoD  
## [1] "0.000911%"  
##   
## $DOptoDis  
## [1] "0.0563%"  
##   
## $MorbProb\_A  
## [1] "100%"  
##   
## $MortProb\_A  
## [1] "100%"  
##   
## $RACE\_NEW  
## [1] "0%"  
##   
## $ETHNICITY\_HISPANIC  
## [1] "0.0000253%"  
##   
## $MORTPROB  
## [1] "1.65%"  
##   
## $MORBPROB  
## [1] "1.65%"  
##   
## $ANESURG  
## [1] "22.3%"  
##   
## $SURGANE  
## [1] "22.3%"  
##   
## $ELECTSURG  
## [1] "17.6%"  
##   
## $DISCHDEST  
## [1] "17.6%"  
##   
## $READMISSION  
## [1] "39.9%"  
##   
## $UNPLANREADMISSION  
## [1] "39.9%"  
##   
## $REOPERATION  
## [1] "39.9%"  
##   
## $STILLINHOSP  
## [1] "17.6%"  
##   
## $REOPERATION1  
## [1] "28.7%"  
##   
## $RETORPODAYS  
## [1] "28.7%"  
##   
## $REOPORCPT1  
## [1] "28.7%"  
##   
## $RETORRELATED  
## [1] "28.7%"  
##   
## $REOPORICD91  
## [1] "28.7%"

BMI is not available in dataset, mutate to combine height/weight

#combine height and weight to make BMI  
nsqip$HEIGHT <- conv\_unit(nsqip$HEIGHT, from = "inch", to= "m")  
nsqip$WEIGHT <- conv\_unit(nsqip$WEIGHT,from ="lbs", to = "kg")  
  
nsqip <- mutate(nsqip, BMI = WEIGHT/HEIGHT^2)  
summary(nsqip$BMI)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -77.34 24.39 28.37 29.48 33.67 9393.01

#remove NAs in MACE  
nsqip$DEATH <- cut(nsqip$DOpertoD, c(-100,0,30))  
levels(nsqip$DEATH) <- c("Alive", "Within 30 Days")

Stratify accordnig to hematocrit

# find percent within hematocrit thresholds (# of patients are columns)  
HCT1 <- nsqip %>% filter(PRHCT <= 24 & PRHCT >0)  
HCT2 <- nsqip %>% filter(PRHCT <= 27 & PRHCT >24)  
HCT3 <- nsqip %>% filter(PRHCT <= 30 & PRHCT >27)  
HCT4 <- nsqip %>% filter(PRHCT <= 33 & PRHCT >30)  
HCT5 <- nsqip %>% filter(PRHCT <= 36 & PRHCT >33)  
HCT6 <- nsqip %>% filter(PRHCT <= 39 & PRHCT >36)  
HCT7 <- nsqip %>% filter(PRHCT >39)

# hematocrit stratification, number of patients in each  
dim(HCT1)

## [1] 27308 115

dim(HCT2)

## [1] 53974 115

dim(HCT3)

## [1] 114741 115

dim(HCT4)

## [1] 206247 115

dim(HCT5)

## [1] 385341 115

dim(HCT6)

## [1] 702218 115

dim(HCT7)

## [1] 1906017 115

#filter patients for target range in preop values  
  
target\_BUN <- nsqip %>% filter(PRBUN >= 5 & PRBUN <= 60)  
target\_CREAT <- nsqip %>% filter(PRCREAT >= 0.02 & PRCREAT <= 15)  
target\_PLT <- nsqip %>% filter(PRPLATE>= 50 & PRPLATE <= 700)  
target\_PTT <- nsqip %>% filter(PRPTT >= 5 & PRPTT <= 25)  
target\_INR <-nsqip %>% filter(PRINR >= 0.1 & PRINR <= 10)  
  
dim(target\_BUN)

## [1] 3037500 115

dim(target\_CREAT)

## [1] 3239597 115

dim(target\_PLT)

## [1] 3318462 115

dim(target\_PTT)

## [1] 175689 115

dim(target\_INR)

## [1] 1703143 115

#number of patients within all target ranges, for reference  
in\_range <- nsqip %>% filter(PRCREAT >= 0.02 & PRCREAT <= 15, PRPLATE>= 50 & PRPLATE <= 700,PRPTT >= 5 & PRPTT <= 25,PRINR >= 0.1 & PRINR <= 10)  
dim(in\_range)

## [1] 166755 115

# Look at days to MACE within 30 days for all patients  
ALLMACE\_30 <- nsqip %>% filter(DOpertoD >0, DCNSCVA >0, DCDMI >0)  
dim(ALLMACE\_30)

## [1] 140 115

CVA\_30 <- nsqip %>% filter(DCNSCVA >0)  
dim(CVA\_30)

## [1] 7029 115

DEATH\_30<- nsqip %>% filter(DOpertoD >0)  
dim(DEATH\_30)

## [1] 41845 115

MI\_30 <-nsqip %>% filter(DCDMI >0)  
dim(MI\_30)

## [1] 12966 115

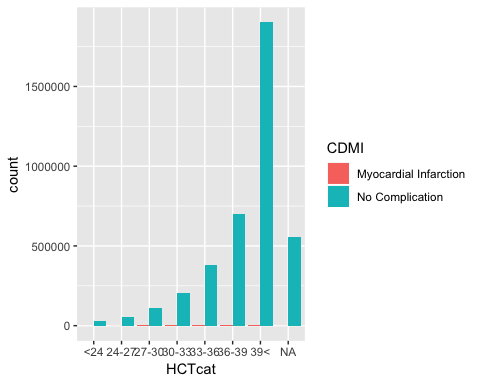
# Look at # of MACE occurences within 30 days for all patients

#NA bool mask  
  
# Remove all NAs to make no\_NA dataframe  
  
#impute NAs in implicit list as 0, no implicit NAs in dataframe  
  
## could not do these steps because of NAs, might need to impute some

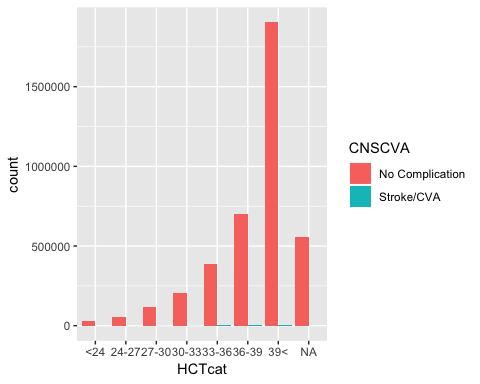
#for each obs, create new factor variable with either occurence of death, MI, or stroke.  
#plot occurence of MIin each HCT category  
nsqip$HCTcat<-cut(nsqip$PRHCT, c(0,24,27,30,33,36,39,300))  
levels(nsqip$HCTcat) <- c("<24", "24-27","27-30","30-33","33-36","36-39","39<")  
summary(nsqip$HCTcat)

## <24 24-27 27-30 30-33 33-36 36-39 39< NA's   
## 27308 53974 114741 206247 385341 702218 1906017 554237

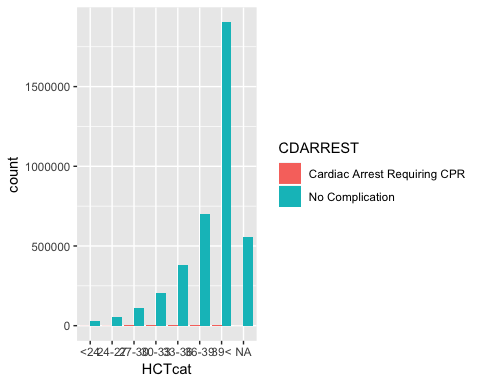
#plots on MACE  
ggplot(data = nsqip) +   
 geom\_bar(mapping = aes(x = HCTcat, fill= CDMI),position = "dodge")



ggplot(data = nsqip) +   
 geom\_bar(mapping = aes(x = HCTcat, fill= CNSCVA),position = "dodge")



ggplot(data = nsqip) +   
 geom\_bar(mapping = aes(x = HCTcat, fill= CDARREST),position = "dodge")



ggplot(data = nsqip) +   
 geom\_bar(mapping = aes(x = HCTcat, fill= DEATH),position = "dodge")

