midcab\_demographics

Salil V Deo

4/18/2020

Report for initial data cleaning for the **midcab** paper.

library(easypackages)  
libraries(c("tidyverse","pastecs","skimr",  
 "readxl","readr","survival","relsurv","lubridate","gmodels"))

## Loading required package: tidyverse

## -- Attaching packages --------- tidyverse 1.3.0 --

## <U+2713> ggplot2 3.2.1 <U+2713> purrr 0.3.3  
## <U+2713> tibble 2.1.3 <U+2713> dplyr 0.8.3  
## <U+2713> tidyr 1.0.0 <U+2713> stringr 1.4.0  
## <U+2713> readr 1.3.1 <U+2713> forcats 0.4.0

## -- Conflicts ------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

## Loading required package: pastecs

##   
## Attaching package: 'pastecs'

## The following objects are masked from 'package:dplyr':  
##   
## first, last

## The following object is masked from 'package:tidyr':  
##   
## extract

## Loading required package: skimr

## Loading required package: readxl

## Loading required package: survival

## Loading required package: relsurv

## Loading required package: date

## Loading required package: lubridate

##   
## Attaching package: 'lubridate'

## The following object is masked from 'package:base':  
##   
## date

## Loading required package: gmodels

## All packages loaded successfully

proc\_desc <- function(x){  
 # require library(pastecs)  
 result <- round(stat.desc(x),2)  
 hist(x, col = "skyblue")  
 result  
}  
  
proc\_tab <- function(x){  
 # require libary(gmodels)  
 CrossTable(x)  
}  
  
  
# na converter  
  
c.na <- function(df,x,y){  
 df$x[df$x == y]<- NA  
}

# get data  
# this the sheet 1 for the file   
  
df <- read\_excel("C:\\github\_rcode\\midcab\\data\\base.xlsx", sheet = 1)

## Warning in read\_fun(path = enc2native(normalizePath(path)), sheet\_i = sheet, :  
## Expecting numeric in FG15 / R15C163: got a date

## Warning in read\_fun(path = enc2native(normalizePath(path)), sheet\_i = sheet, :  
## Expecting numeric in FJ644 / R644C166: got a date

## Warning in read\_fun(path = enc2native(normalizePath(path)), sheet\_i = sheet, :  
## Expecting numeric in FJ1626 / R1626C166: got a date

## Warning in read\_fun(path = enc2native(normalizePath(path)), sheet\_i =  
## sheet, : Expecting numeric in C1809 / R1809C3: got '7809224773, neue nummer:  
## 015906188975'

## Warning in read\_fun(path = enc2native(normalizePath(path)), sheet\_i = sheet, :  
## Expecting logical in FL1884 / R1884C168: got 'Herzklappenverengung, Op wurde auf  
## Pat.wunschabgsagt'

## Warning in read\_fun(path = enc2native(normalizePath(path)), sheet\_i = sheet, :  
## Expecting logical in FL1893 / R1893C168: got 'Herzinf. 04/2018, KÃ¶rper reagiert  
## gegen Stenmaterial, evtl. Chronifizierung'

## Warning in read\_fun(path = enc2native(normalizePath(path)), sheet\_i = sheet, :  
## Expecting logical in FL2007 / R2007C168: got '3 Stents, 2018'

## Warning in read\_fun(path = enc2native(normalizePath(path)), sheet\_i = sheet, :  
## Expecting logical in FL2012 / R2012C168: got 'Stent nach Hinf., 12/2018'

## Warning in read\_fun(path = enc2native(normalizePath(path)), sheet\_i = sheet, :  
## Expecting logical in FL2266 / R2266C168: got 'Mai 2019, 2 Stents'

## Warning in read\_fun(path = enc2native(normalizePath(path)), sheet\_i = sheet, :  
## Expecting numeric in FJ2283 / R2283C166: got a date

dim(df)

## [1] 2450 196

glimpse(df)

## Observations: 2,450  
## Variables: 196  
## $ lfdNr <dbl> 1, 2, 3, 5, 6, 7, 8…  
## $ THG\_Referenz <chr> "ACB 4-fach mit HLM…  
## $ Aufnahme\_Nr <dbl> 7800161501, 7800499…  
## $ Pat\_ID <dbl> 78001615, 78004997,…  
## $ date\_of\_surgery <dttm> 1995-05-22, 1996-0…  
## $ dob <dttm> 1948-08-22, 1948-1…  
## $ age\_at\_surgery <dbl> 46.75, 47.44, 67.48…  
## $ gender <dbl> 1, 2, 2, 1, 1, 1, 1…  
## $ admission <dttm> 1995-05-18, 1996-0…  
## $ height\_cm <dbl> -9, 159, 156, -9, 1…  
## $ weight\_kg <dbl> -9, 68, 63, 92, 93,…  
## $ BMI <dbl> -9.0, 26.9, 25.9, -…  
## $ BSA\_Mosteller <dbl> -9.000, 1.733, 1.65…  
## $ pre\_med\_ass <dbl> -9, 1, 0, -9, 0, 1,…  
## $ pre\_med\_clopidogrel <dbl> -9, -9, -9, -9, -9,…  
## $ pre\_med\_Gp2b\_3A\_antagonist <dbl> -9, -9, -9, -9, -9,…  
## $ pre\_med\_nitrate\_iv <dbl> 0, -9, 0, -9, -9, -…  
## $ pre\_med\_inotropika <dbl> 0, -9, 0, -9, -9, -…  
## $ diabetes <dbl> 0, 0, 0, 0, -9, 0, …  
## $ art\_hypertension <dbl> -9, 1, 0, -9, 1, 0,…  
## $ `pulm\_hypertension>60mmHg` <dbl> 0, -9, 0, -9, -9, -…  
## $ smoker <dbl> -9, -9, 0, -9, -9, …  
## $ hyperlipemia <dbl> -9, 0, 0, -9, 1, 1,…  
## $ COPD <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ peripheral\_vascular\_disease <dbl> 0, 0, 0, -9, 0, 0, …  
## $ pre\_creatinine\_mg\_dl <dbl> -9, -9, -9, -9, -9,…  
## $ pre\_dialysis <dbl> 0, 0, 0, -9, 0, 0, …  
## $ preGFR\_CockcroftGault <dbl> -9, -9, -9, -9, -9,…  
## $ pre\_LVEF <dbl> 46, -9, 69, -9, 72,…  
## $ pre\_LVEF\_function <dbl> 2, 1, 1, -9, 1, 1, …  
## $ pre\_heartrhythm <dbl> 2, 1, 2, -9, 1, 2, …  
## $ PM\_wearer <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ ICD\_wearer <dbl> 0, -9, 0, -9, -9, -…  
## $ pre\_NYHA <dbl> -9, -9, 3, -9, -9, …  
## $ pre\_CCS <dbl> -9, -9, 3, -9, -9, …  
## $ pre\_ASA <dbl> -9, -9, 3, -9, -9, …  
## $ priority <dbl> 1, 1, 1, 1, 2, 1, 1…  
## $ elective <dbl> 1, 1, 1, 1, 0, 1, 1…  
## $ urgent <dbl> 0, 0, 0, 0, 1, 0, 0…  
## $ emergency <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ critical\_preoperative\_state <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ pre\_reanimation <dbl> 0, -9, 0, -9, -9, -…  
## $ pre\_ventilation <dbl> 0, -9, 0, -9, -9, -…  
## $ coronary\_disease\_detail <dbl> 0, -9, 0, -9, -9, -…  
## $ left\_main\_disease <dbl> -9, -9, -9, -9, -9,…  
## $ pre\_lyse <lgl> NA, NA, NA, NA, NA,…  
## $ prior\_PCI <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ last\_PCI\_date <dttm> NA, NA, NA, NA, NA…  
## $ pre\_MI <dbl> 0, -9, 0, -9, 0, -9…  
## $ `pre\_MI<48hours` <dbl> 0, 0, 0, -9, 0, 0, …  
## $ `pre\_MI\_2\_to\_<21days` <dbl> 0, 0, 0, -9, 0, 0, …  
## $ `pre\_MI\_21\_to\_<91days` <dbl> 0, 0, 0, -9, 0, 0, …  
## $ `pre\_MI\_>90days` <dbl> 0, 0, 0, -9, 0, 0, …  
## $ pre\_CVA <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ pre\_neurological\_dysfunction <dbl> 0, 0, 0, -9, 0, 0, …  
## $ pre\_cardiac\_shock <dbl> 0, 0, 0, -9, 0, 1, …  
## $ `pre\_cardiac\_shock<48hours` <dbl> 0, 0, 0, -9, 0, 0, …  
## $ pre\_iabp <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ pre\_ecmo <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ typ\_a\_dissection <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ diss\_iatrogen <lgl> NA, NA, NA, NA, NA,…  
## $ active\_endocarditis <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ postinfarct\_septal\_rupture <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ prior\_cardiac\_surgery <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ last\_cardiac\_surgery\_date <dttm> NA, NA, NA, NA, NA…  
## $ additive\_euroSCORE <dbl> 1, 0, 2, 0, 0, 3, 0…  
## $ logistic\_euroscore <dbl> 1.33, 1.22, 2.20, 0…  
## $ CABG <dbl> 1, 1, 1, 1, 1, 1, 1…  
## $ OPCAB <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ MIDCAB <dbl> 0, 1, 1, 1, 1, 1, 1…  
## $ TECAB <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ conversion\_to\_onpump <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ onpump\_beating\_heart <dbl> 0, 1, 0, 1, 1, 1, 0…  
## $ incompl\_revascularisation <dbl> -9, 0, 0, 0, 0, 0, …  
## $ no\_touch\_aorta <dbl> -9, -9, -9, -9, -9,…  
## $ AV\_replacement <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ AV\_prosthesis\_type <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ AV\_prosthesis\_model <lgl> NA, NA, NA, NA, NA,…  
## $ AV\_prosthesis\_diam <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ AV\_Conduit <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ AV\_repair <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ MV\_replacement <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ MV\_replacement\_1stfailed <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ MV\_prosthesis\_type <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ MV\_prosthesis\_model <lgl> NA, NA, NA, NA, NA,…  
## $ MV\_prosthesis\_diam <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ MV\_repair <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ MV\_repair\_failed <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ TV\_replacement <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ TV\_prosthesis\_type <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ TV\_prosthesis\_model <lgl> NA, NA, NA, NA, NA,…  
## $ TV\_prosthesis\_diam <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ TV\_repair <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ morrow\_resection <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ Maze <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ ASD <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ cardiac\_tumor\_resection <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ AAE <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ ABE <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ arch\_type <lgl> NA, NA, NA, NA, NA,…  
## $ arch\_remark <lgl> NA, NA, NA, NA, NA,…  
## $ arch\_elephant\_trunk <lgl> NA, NA, NA, NA, NA,…  
## $ arch\_elephant\_trunk\_prosthesis <lgl> NA, NA, NA, NA, NA,…  
## $ initial\_thoracotomy <dbl> 4, -9, 3, 3, 3, 3, …  
## $ conversion\_to\_full\_sternotomy <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ length\_of\_surgery <dbl> 0, 170, 145, 114, 1…  
## $ bypass\_time <dbl> 92, 42, 38, 31, 26,…  
## $ cross\_clamp\_time <dbl> 57, 0, 24, 0, 0, 0,…  
## $ cardiopleg\_type <dbl> 1, 0, 1, 0, 0, 0, 1…  
## $ cardiopleg\_tech <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ cardiopleg\_ml <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ min\_temperature <dbl> 0, 32, 33, 36, 35, …  
## $ post\_low\_cardiac\_output <dbl> 0, 0, 0, -9, 0, -9,…  
## $ post\_iabp <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ post\_ecmo <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ post\_MI <dbl> 0, 0, 0, -9, 0, -9,…  
## $ post\_reanimation <dbl> 0, 0, 0, -9, 0, -9,…  
## $ post\_cardiac\_arrhythmia <dbl> -9, 0, 0, -9, 1, -9…  
## $ post\_redo\_heart <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ post\_rethorax\_bleeding <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ post\_symptomatic\_transitory\_psychotic\_syndrome <dbl> 0, 0, 0, -9, 0, -9,…  
## $ post\_CVA <dbl> 0, -9, -9, -9, -9, …  
## $ `post\_CVA>72hours` <lgl> NA, NA, NA, NA, NA,…  
## $ `post\_CVA\_rankin>0` <lgl> NA, NA, NA, NA, NA,…  
## $ post\_DSWI <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ post\_sepsis <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ post\_GI\_complications <dbl> 0, 0, 0, -9, 0, -9,…  
## $ post\_maxCreatinine\_mg\_dl <dbl> -9, -9, -9, -9, -9,…  
## $ post\_maxGFR\_CockcroftGault <dbl> -9, -9, -9, -9, -9,…  
## $ post\_dialysis <dbl> 0, 0, 0, -9, 0, -9,…  
## $ EK <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ TK <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ FFP <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ ventilation\_time\_until\_first\_extubation <dbl> -9, -9, -9, -9, -9,…  
## $ ventilation\_time\_total <dbl> -9, -9, -9, -9, -9,…  
## $ post\_respiratory\_failure <dbl> 0, 0, 0, -9, 0, -9,…  
## $ re\_intubation <dbl> NA, NA, NA, NA, NA,…  
## $ tracheotomie <dbl> 0, 0, 0, -9, 0, -9,…  
## $ post\_heartrhythm <dbl> 1, 1, 1, -9, 1, -9,…  
## $ post\_new\_PM <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ post\_new\_ICD <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ dayspost <dbl> 10, 7, 18, 7, 11, 9…  
## $ ICU\_1stLOS <lgl> NA, NA, NA, NA, NA,…  
## $ ICU\_LOS <lgl> NA, NA, NA, NA, NA,…  
## $ n\_ICU <lgl> NA, NA, NA, NA, NA,…  
## $ Re\_ICU <lgl> NA, NA, NA, NA, NA,…  
## $ IMCU\_1stLOS <lgl> NA, NA, NA, NA, NA,…  
## $ IMCU\_LOS <lgl> NA, NA, NA, NA, NA,…  
## $ n\_IMCU <lgl> NA, NA, NA, NA, NA,…  
## $ Re\_IMCU <lgl> NA, NA, NA, NA, NA,…  
## $ discharge <dttm> 1995-06-01, 1996-0…  
## $ died <dbl> 1, 0, 0, 0, 0, 0, 0…  
## $ died\_inhouse <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ died\_fu <dbl> 1, 0, 0, 0, 0, 0, 0…  
## $ `survival\_days=` <dbl> 1610, 5660, 5611, 5…  
## $ last\_contact\_date <dttm> 1999-10-28, 2011-1…  
## $ S\_Datum <chr> "36461", NA, NA, NA…  
## $ Todesursache <chr> "unbekannt", NA, NA…  
## $ ReOP1\_Date <chr> NA, NA, NA, NA, NA,…  
## $ ReOP1\_Text <chr> NA, NA, NA, NA, NA,…  
## $ ReOP2\_Date <chr> NA, NA, NA, NA, NA,…  
## $ Schlaganfall <chr> NA, NA, NA, NA, "0"…  
## $ Herzinfarkt <dbl> NA, NA, NA, NA, 0.0…  
## $ `Revask.LAD bereich 1=AcB, 2=PTCA,Stent 0=nichts` <chr> NA, NA, NA, NA, "0"…  
## $ `AP CCS` <dbl> NA, NA, NA, NA, 1, …  
## $ Revaskul.nichtLAD <dbl> NA, NA, NA, NA, NA,…  
## $ `Date nicht LAD` <chr> NA, NA, NA, NA, NA,…  
## $ ReOP5\_Text <lgl> NA, NA, NA, NA, NA,…  
## $ lad\_area <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ rcx\_area <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ rca\_area <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ Graft1 <dbl> 3, 3, 3, 3, 3, 3, 3…  
## $ Graft2 <dbl> 9, NA, NA, NA, NA, …  
## $ Graft3 <dbl> 9, NA, NA, NA, NA, …  
## $ Graft4 <dbl> 9, NA, NA, NA, NA, …  
## $ Graft5 <lgl> NA, NA, NA, NA, NA,…  
## $ LIMA <dbl> 1, 1, 1, 1, 1, 1, 1…  
## $ RIMA <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ BIMA <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ Radialis <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ Vene <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ sonstGraft <dbl> 1, 0, 0, 0, 0, 0, 0…  
## $ Anz\_Jump\_Seq\_Anastomosen <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ Anz\_aortaleAnastomosen <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ Anz\_YT\_Anastomosen <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ LIMA\_proximal <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ RIMA\_proximal <lgl> NA, NA, NA, NA, NA,…  
## $ Radialis\_proximal <lgl> NA, NA, NA, NA, NA,…  
## $ anzahlArteriellerGrafts <dbl> 1, 1, 1, 1, 1, 1, 1…  
## $ lima\_lad <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ lima\_freegraft <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ rima\_freegraft <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ lad\_graft <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ rcx\_graft <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ rca\_graft <dbl> 0, 0, 0, 0, 0, 0, 0…  
## $ Anzahl\_Grafts <dbl> 4, 1, 1, 1, 1, 1, 1…

# create data into sections according to the information  
# Pat\_ID is the patient indicator  
# create table demo with baseline demographics first and then clean that  
  
  
demo <- c("lfdNr", "THG\_Referenz", "Aufnahme\_Nr", "Pat\_ID", "date\_of\_surgery",   
 "dob", "age\_at\_surgery", "gender", "admission", "height\_cm",   
 "weight\_kg", "BMI", "BSA\_Mosteller", "pre\_med\_ass", "pre\_med\_clopidogrel",   
 "pre\_med\_Gp2b\_3A\_antagonist", "pre\_med\_nitrate\_iv", "pre\_med\_inotropika",   
 "diabetes", "art\_hypertension", "pulm\_hypertension>60mmHg", "smoker",   
 "hyperlipemia", "COPD", "peripheral\_vascular\_disease", "pre\_creatinine\_mg\_dl",   
 "pre\_dialysis", "preGFR\_CockcroftGault", "pre\_LVEF", "pre\_LVEF\_function",   
 "pre\_heartrhythm", "PM\_wearer", "ICD\_wearer", "pre\_NYHA", "pre\_CCS",   
 "pre\_ASA", "priority", "elective", "urgent", "emergency", "critical\_preoperative\_state",   
 "pre\_reanimation", "pre\_ventilation", "coronary\_disease\_detail",   
 "left\_main\_disease", "pre\_lyse", "prior\_PCI", "last\_PCI\_date",   
 "pre\_MI", "pre\_MI<48hours", "pre\_MI\_2\_to\_<21days", "pre\_MI\_21\_to\_<91days",   
 "pre\_MI\_>90days", "pre\_CVA", "pre\_neurological\_dysfunction",   
 "pre\_cardiac\_shock", "pre\_cardiac\_shock<48hours", "pre\_iabp",   
 "pre\_ecmo", "typ\_a\_dissection", "diss\_iatrogen", "active\_endocarditis",   
 "postinfarct\_septal\_rupture", "prior\_cardiac\_surgery", "last\_cardiac\_surgery\_date",   
 "additive\_euroSCORE", "logistic\_euroscore", "CABG", "OPCAB",   
 "MIDCAB", "TECAB")  
  
demog <- df[,c(demo)] # demog now has all the demographic information for all the patients  
  
# check if the rows are unique  
  
duplicates <- demog[duplicated(demog$Pat\_ID),] # duplicated numbers   
  
# duplicate patient id  
  
duplicates %>% select(Pat\_ID)

## # A tibble: 1 x 1  
## Pat\_ID  
## <dbl>  
## 1 78071802

There is 1 duplicate patient. This is the demographics table; all that is the same. So I have kept 1 row and removed the other in table *demog2*.

demog2 <- demog %>%   
 group\_by(Pat\_ID) %>%  
 filter(!duplicated(Pat\_ID)| n() == 1)

To confirm that all unique in table *demog2*

duplicates <- demog2[duplicated(demog2$Pat\_ID), ]  
  
dim(duplicates)

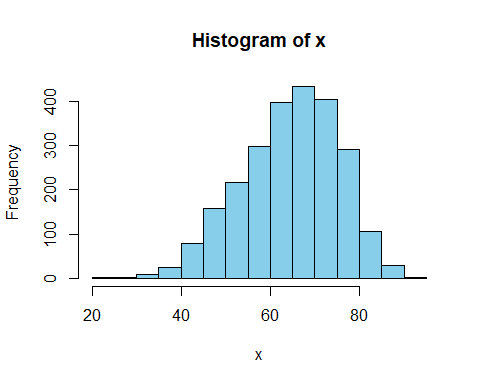
## [1] 0 71

Now to see the data and work on cleaning. *demog2* now has 71 col. That is going to be the master df now.

# get the first 10 col  
  
d <- demog2[, 1:10]  
  
glimpse(d)

## Observations: 2,449  
## Variables: 10  
## Groups: Pat\_ID [2,449]  
## $ lfdNr <dbl> 1, 2, 3, 5, 6, 7, 8, 9, 10, 13, 14, 15, 17, 18, 19, 2…  
## $ THG\_Referenz <chr> "ACB 4-fach mit HLM", "ACB 1-fach MIDCAB mit HLM", "A…  
## $ Aufnahme\_Nr <dbl> 7800161501, 7800499704, 7800825702, 7800969201, 78009…  
## $ Pat\_ID <dbl> 78001615, 78004997, 78008257, 78009692, 78009755, 780…  
## $ date\_of\_surgery <dttm> 1995-05-22, 1996-05-03, 1996-05-03, 1996-05-15, 1996…  
## $ dob <dttm> 1948-08-22, 1948-11-24, 1928-11-11, 1937-12-01, 1944…  
## $ age\_at\_surgery <dbl> 46.75, 47.44, 67.48, 58.46, 52.18, 74.33, 50.17, 65.7…  
## $ gender <dbl> 1, 2, 2, 1, 1, 1, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,…  
## $ admission <dttm> 1995-05-18, 1996-04-30, 1996-04-26, 1996-05-13, 1996…  
## $ height\_cm <dbl> -9, 159, 156, -9, 172, 180, 178, 168, 178, 172, 164, …

# age   
  
proc\_desc(d$age\_at\_surgery)



## nbr.val nbr.null nbr.na min max range   
## 2449.00 0.00 0.00 24.69 92.41 67.72   
## sum median mean SE.mean CI.mean.0.95 var   
## 158125.33 65.56 64.57 0.22 0.43 118.91   
## std.dev coef.var   
## 10.90 0.17

# age ok ; no missing   
  
# convert dates to date in R   
  
d$date\_of\_surgery <- as\_date(d$date\_of\_surgery)  
  
d$dob <- as\_date(d$dob)  
  
# gender  
  
str(d$gender)

## num [1:2449] 1 2 2 1 1 1 1 2 1 1 ...

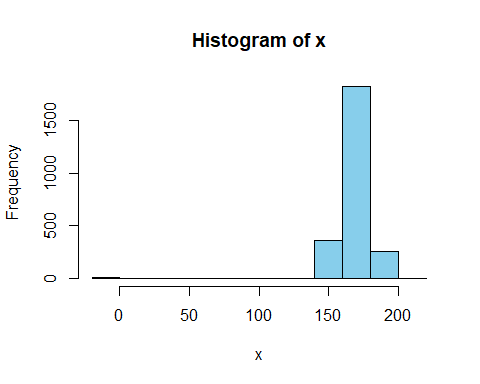
table(d$gender, useNA = 'ifany') # no missing gender 1 = male, 2= female

##   
## 1 2   
## 1818 631

# conv to female  
  
d$female <- d$gender - 1  
  
d$female <- factor(d$female, levels = c(0,1), labels = c("no","yes"))  
  
proc\_tab(d$female)

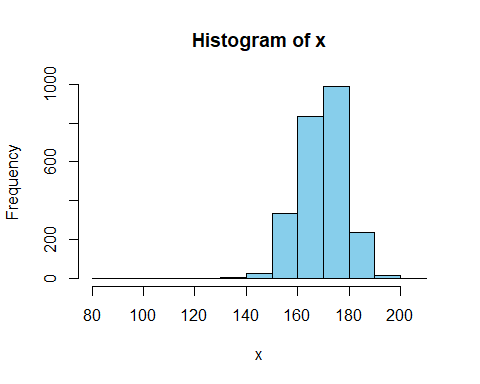
##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | N / Table Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 2449   
##   
##   
## | no | yes |   
## |-----------|-----------|  
## | 1818 | 631 |   
## | 0.742 | 0.258 |   
## |-----------|-----------|  
##   
##   
##   
##

# height\_cm  
  
proc\_desc(d$height\_cm) # missing = -9



## nbr.val nbr.null nbr.na min max range   
## 2449.00 0.00 0.00 -9.00 205.00 214.00   
## sum median mean SE.mean CI.mean.0.95 var   
## 415711.00 171.00 169.75 0.29 0.58 212.91   
## std.dev coef.var   
## 14.59 0.09

# convert that to missing data  
  
d$height\_cm[d$height\_cm == -9]<- NA  
  
proc\_desc(d$height\_cm)



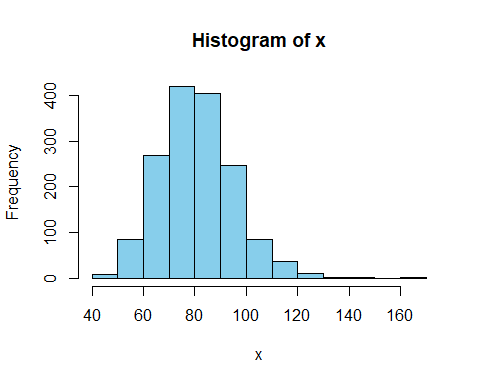
## nbr.val nbr.null nbr.na min max range   
## 2439.00 0.00 10.00 88.00 205.00 117.00   
## sum median mean SE.mean CI.mean.0.95 var   
## 415801.00 171.00 170.48 0.18 0.36 82.20   
## std.dev coef.var   
## 9.07 0.05

now go for the next 10 variables

d2 <- demog2 %>% select(Pat\_ID,11:20)  
  
glimpse(d2)

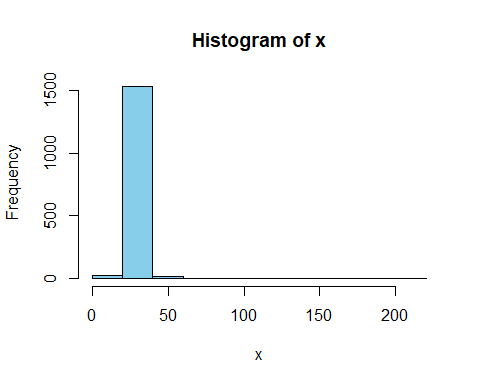
## Observations: 2,449  
## Variables: 11  
## Groups: Pat\_ID [2,449]  
## $ Pat\_ID <dbl> 78001615, 78004997, 78008257, 78009692, 78…  
## $ weight\_kg <dbl> -9, 68, 63, 92, 93, 78, 82, 67, 80, 84, 66…  
## $ BMI <dbl> -9.0, 26.9, 25.9, -9.0, 31.4, 24.1, 25.9, …  
## $ BSA\_Mosteller <dbl> -9.000, 1.733, 1.652, -9.000, 2.108, 1.975…  
## $ pre\_med\_ass <dbl> -9, 1, 0, -9, 0, 1, 0, 0, 1, 1, 0, 0, 0, 1…  
## $ pre\_med\_clopidogrel <dbl> -9, -9, -9, -9, -9, -9, -9, -9, -9, -9, -9…  
## $ pre\_med\_Gp2b\_3A\_antagonist <dbl> -9, -9, -9, -9, -9, -9, -9, -9, -9, -9, -9…  
## $ pre\_med\_nitrate\_iv <dbl> 0, -9, 0, -9, -9, -9, -9, -9, -9, -9, -9, …  
## $ pre\_med\_inotropika <dbl> 0, -9, 0, -9, -9, -9, -9, -9, -9, -9, -9, …  
## $ diabetes <dbl> 0, 0, 0, 0, -9, 0, 0, -9, 0, 0, 0, 0, 0, -…  
## $ art\_hypertension <dbl> -9, 1, 0, -9, 1, 0, 0, 1, 1, 1, 0, 0, 1, 1…

# weight  
  
d2$weight\_kg[d2$weight\_kg == -9]<- NA  
  
proc\_desc(d2$weight\_kg)



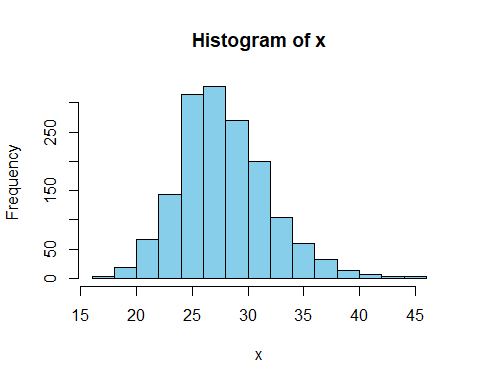
## nbr.val nbr.null nbr.na min max range   
## 1572.00 0.00 877.00 44.00 168.00 124.00   
## sum median mean SE.mean CI.mean.0.95 var   
## 128551.00 81.00 81.78 0.37 0.72 213.82   
## std.dev coef.var   
## 14.62 0.18

# BMI  
  
d2$BMI[d2$BMI == -9]<- NA  
  
proc\_desc(d2$BMI)



## nbr.val nbr.null nbr.na min max range   
## 1569.00 0.00 880.00 16.10 216.90 200.80   
## sum median mean SE.mean CI.mean.0.95 var   
## 43984.30 27.50 28.03 0.16 0.31 39.64   
## std.dev coef.var   
## 6.30 0.22

# make BMI meaningful so if BMI > 45, then converted to 45  
  
d2$BMI[d2$BMI > 45]<- 45  
  
proc\_desc(d2$BMI)



## nbr.val nbr.null nbr.na min max range   
## 1569.00 0.00 880.00 16.10 45.00 28.90   
## sum median mean SE.mean CI.mean.0.95 var   
## 43805.10 27.50 27.92 0.10 0.20 16.88   
## std.dev coef.var   
## 4.11 0.15

# antiplatelet agent before surgery  
# either aspirin, clopidogrel or other G2P3 antagonist  
  
d2$antiplt <- with(d2, ifelse((pre\_med\_ass == 1|pre\_med\_clopidogrel == 1| pre\_med\_Gp2b\_3A\_antagonist == 1), 1, 0))  
  
proc\_tab(d2$antiplt)

##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | N / Table Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 2449   
##   
##   
## | 0 | 1 |   
## |-----------|-----------|  
## | 878 | 1571 |   
## | 0.359 | 0.641 |   
## |-----------|-----------|  
##   
##   
##   
##

# diabetes  
  
proc\_tab(d2$diabetes)

##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | N / Table Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 2449   
##   
##   
## | -9 | 0 | 1 |   
## |-----------|-----------|-----------|  
## | 273 | 1689 | 487 |   
## | 0.111 | 0.690 | 0.199 |   
## |-----------|-----------|-----------|  
##   
##   
##   
##

d2$diabetes[d2$diabetes == -9]<- NA  
  
proc\_tab(d2$diabetes)

##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | N / Table Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 2176   
##   
##   
## | 0 | 1 |   
## |-----------|-----------|  
## | 1689 | 487 |   
## | 0.776 | 0.224 |   
## |-----------|-----------|  
##   
##   
##   
##

now to tackle col 21:30 of the data

d3 <- demog2 %>% select(Pat\_ID, 21:30)