

Appendix - C - R Markdown File

BSc Population Health Dissertation (18/19)

```
##### Package Library #####  
#####
```

```
library(data.table)
```

```
## Warning: package 'data.table' was built under R version 3.4.4
```

```
library(readr)
```

```
## Warning: package 'readr' was built under R version 3.4.4
```

```
library(foreign)
```

```
## Warning: package 'foreign' was built under R version 3.4.4
```

```
library(haven)
```

```
## Warning: package 'haven' was built under R version 3.4.4
```

```
library(grid)  
library(gridExtra)  
library(factoextra)
```

```
## Loading required package: ggplot2
```

```
## Warning: package 'ggplot2' was built under R version 3.4.4
```

```
## Welcome! Related Books: `Practical Guide To Cluster Analysis in R` at https://goo.gl/13EFCZ
```

```
library(mice)
```

```
## Warning: package 'mice' was built under R version 3.4.4
```

```
## Loading required package: lattice
```

```
## Warning: package 'lattice' was built under R version 3.4.4
```

```
##  
## Attaching package: 'mice'
```

```
## The following objects are masked from 'package:base':  
##  
##      cbind, rbind
```

```
library(PerformanceAnalytics)
```

```
## Loading required package: xts
```

```
## Warning: package 'xts' was built under R version 3.4.4
```

```
## Loading required package: zoo
```

```
## Warning: package 'zoo' was built under R version 3.4.4
```

```
##  
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':  
##  
##      as.Date, as.Date.numeric
```

```
##  
## Attaching package: 'xts'
```

```
## The following objects are masked from 'package:data.table':  
##  
##      first, last
```

```
##  
## Attaching package: 'PerformanceAnalytics'
```

```
## The following object is masked from 'package:graphics':  
##  
##      legend
```

```
library(factoextra)  
library(ca)
```

```
## Warning: package 'ca' was built under R version 3.4.4
```

```
library(highcharter)
```

```
## Highcharts (www.highcharts.com) is a Highsoft software product which is
```

```
## not free for commercial and Governmental use
```

```
library(tidyverse)
```

```
## — Attaching packages ————— tidyverse 1.2.1 —
```

```
## ✓ tibble 1.4.2      ✓ dplyr 0.7.8
## ✓ tidyr 0.8.2      ✓ stringr 1.3.1
## ✓ purrr 0.2.5      ✓ forcats 0.3.0
```

```
## Warning: package 'tidyr' was built under R version 3.4.4
```

```
## Warning: package 'purrr' was built under R version 3.4.4
```

```
## Warning: package 'dplyr' was built under R version 3.4.4
```

```
## Warning: package 'stringr' was built under R version 3.4.4
```

```
## — Conflicts ————— tidyverse_conflicts() —
## ✗ dplyr::between() masks data.table::between()
## ✗ dplyr::combine() masks gridExtra::combine()
## ✗ tidyr::complete() masks mice::complete()
## ✗ dplyr::filter() masks stats::filter()
## ✗ dplyr::first() masks xts::first(), data.table::first()
## ✗ dplyr::lag() masks stats::lag()
## ✗ dplyr::last() masks xts::last(), data.table::last()
## ✗ purrr::transpose() masks data.table::transpose()
```

```
library(rwars)
library(Matching)
```

```
## Warning: package 'Matching' was built under R version 3.4.4
```

```
## Loading required package: MASS
```

```
## Warning: package 'MASS' was built under R version 3.4.4
```

```
##
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':  
##  
##   select
```

```
## ##  
## ## Matching (Version 4.9-3, Build Date: 2018-05-03)  
## ## See http://sekhon.berkeley.edu/matching for additional documentation.  
## ## Please cite software as:  
## ##   Jasjeet S. Sekhon. 2011. ``Multivariate and Propensity Score Matching  
## ##   Software with Automated Balance Optimization: The Matching package for R.''  
## ##   Journal of Statistical Software, 42(7): 1-52.  
## ##
```

```
library(Hmisc)
```

```
## Loading required package: survival
```

```
## Warning: package 'survival' was built under R version 3.4.4
```

```
## Loading required package: Formula
```

```
## Warning: package 'Formula' was built under R version 3.4.4
```

```
##  
## Attaching package: 'Hmisc'
```

```
## The following objects are masked from 'package:dplyr':  
##  
##   src, summarize
```

```
## The following objects are masked from 'package:base':  
##  
##   format.pval, units
```

```
library(dplyr)  
library(tableone)
```

```
## Warning: package 'tableone' was built under R version 3.4.4
```

```
library(stargazer)
```

```
## Warning: package 'stargazer' was built under R version 3.4.4
```

```
##  
## Please cite as:
```

```
## Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics  
Tables.
```

```
## R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
```

```
library(Amelia)
```

```
## Warning: package 'Amelia' was built under R version 3.4.4
```

```
## Loading required package: Rcpp
```

```
## Warning: package 'Rcpp' was built under R version 3.4.4
```

```
## ##  
## ## Amelia II: Multiple Imputation  
## ## (Version 1.7.5, built: 2018-05-07)  
## ## Copyright (C) 2005-2019 James Honaker, Gary King and Matthew Blackwell  
## ## Refer to http://gking.harvard.edu/amelia/ for more information  
## ##
```

```
library(cobalt)
```

```
## Warning: package 'cobalt' was built under R version 3.4.4
```

```
##### Write in 2005-2015 datasets #####  
#####  
  
hse05 = read_dta("/Users/vincentmay/Desktop/Dissertation/Codes/dataset/hse05ai.dta")  
  
hse06 = read_dta("/Users/vincentmay/Desktop/Dissertation/Codes/dataset/hse06ai.dta")  
  
hse07 = read_dta("/Users/vincentmay/Desktop/Dissertation/Codes/dataset/hse07ai.dta")  
  
hse08 = read_dta("/Users/vincentmay/Desktop/Dissertation/Codes/dataset/hse08ai.dta")  
  
hse09 = read_dta("/Users/vincentmay/Desktop/Dissertation/Codes/dataset/hse09ai.dta")  
  
hse10 = read_dta("/Users/vincentmay/Desktop/Dissertation/Codes/dataset/hse10ai.dta")  
  
hse11 = read_dta("/Users/vincentmay/Desktop/Dissertation/Codes/dataset/hse2011ai.dta"  
)  
  
hse12 = read_dta("/Users/vincentmay/Desktop/Dissertation/Codes/dataset/hse2012ai.dta"  
)  
  
hse13 = read_dta("/Users/vincentmay/Desktop/Dissertation/Codes/dataset/hse2013ai.dta"  
)  
  
hse14 = read_dta("/Users/vincentmay/Desktop/Dissertation/Codes/dataset/hse2014ai.dta"  
)  
  
hse15 = read_dta("/Users/vincentmay/Desktop/Dissertation/Codes/dataset/hse2015ai.dta"  
)
```

Part 1/3: Data Preparation

```
##### Data Preparation -- Get variables needed
#####

# In this section, variables in interest are selected.
# Children are filtered from the datasets by "age", which is further converted to three age groups

hse05.pc <- c("syslom", "diaslom", "sex", "tenureb", "age", "ethinda", "hhsz", "addnum", "imd2004",
             "birthwt", "porftvg", "gor", "sys2om", "sys3om", "dias2om", "dias3om")
hse05.mk2 <- hse05[,hse05.pc]
colnames(hse05.mk2)[6] <- "origin"
colnames(hse05.mk2)[9] <- "imd"
hse05.mk2$year <- 05
hse05.mk2$aggr <- ifelse(hse05.mk2$age > 10 & hse05.mk2$age < 16 , 3,
                        ifelse(hse05.mk2$age > 4 & hse05.mk2$age < 11 , 2,
                              ifelse(hse05.mk2$age > 1 & hse05.mk2$age < 5 , 1, 0
)))
hse05.mk2$porftvg <- ifelse(hse05.mk2$porftvg > 5, 6, hse05.mk2$porftvg)
hse05.2 <- -c(which(hse05.mk2$aggr == 0))
hse05.mk2 <- hse05.mk2[hse05.2,]
hse05.mk2$age <- NULL

hse06.pc <- c("syslom", "diaslom", "sex", "tenureb", "age", "ethinda", "hhsz", "addnum", "imd2004",
             "birthwt", "porftvg", "gor06", "sys2om", "sys3om", "dias2om", "dias3om")
hse06.mk2 <- hse06[,hse06.pc]
colnames(hse06.mk2)[6] <- "origin"
colnames(hse06.mk2)[9] <- "imd"
colnames(hse06.mk2)[12] <- "gor"
hse06.mk2$year <- 06
hse06.mk2$aggr <- ifelse(hse06.mk2$age > 10 & hse06.mk2$age < 16 , 3,
                        ifelse(hse06.mk2$age > 4 & hse06.mk2$age < 11 , 2,
                              ifelse(hse06.mk2$age > 1 & hse06.mk2$age < 5 , 1, 0
)))
hse06.mk2$porftvg <- ifelse(hse06.mk2$porftvg > 5, 6, hse06.mk2$porftvg)
hse06.2 <- -c(which(hse06.mk2$aggr == 0))
hse06.mk2 <- hse06.mk2[hse06.2,]
hse06.mk2$age <- NULL

hse07.pc <- c("syslom", "diaslom", "sex", "tenureb", "age", "ethinda", "hhsz", "addnum", "imd2007",
             "birthwt", "porftvg", "gor07", "sys2om", "sys3om", "dias2om", "dias3om")
hse07.mk2 <- hse07[,hse07.pc]
colnames(hse07.mk2)[6] <- "origin"
colnames(hse07.mk2)[7] <- "hhsz"
colnames(hse07.mk2)[9] <- "imd"
colnames(hse07.mk2)[12] <- "gor"
hse07.mk2$year <- 07
hse07.mk2$aggr <- ifelse(hse07.mk2$age > 10 & hse07.mk2$age < 16 , 3,
                        ifelse(hse07.mk2$age > 4 & hse07.mk2$age < 11 , 2,
                              ifelse(hse07.mk2$age > 1 & hse07.mk2$age < 5 , 1, 0
)))
hse07.mk2$porftvg <- ifelse(hse07.mk2$porftvg > 5, 6, hse07.mk2$porftvg)
hse07.2 <- -c(which(hse07.mk2$aggr == 0))
```

```

hse07.mk2 <- hse07.mk2[hse07.2,]
hse07.mk2$age <- NULL

hse08.pc <- c("syslom","diaslom","sex","tenureb","age","origin","hhsz","addnum","qim",
             "birthwt","porftvg","GOR","sys2om","sys3om","dias2om","dias3om")
hse08.mk2 <- hse08[,hse08.pc]
colnames(hse08.mk2)[9] <- "imd"
colnames(hse08.mk2)[12] <- "gor"
hse08.mk2$year <- 08
hse08.mk2$aggr <- ifelse(hse08.mk2$age > 10 & hse08.mk2$age < 16 , 3,
                        ifelse(hse08.mk2$age > 4 & hse08.mk2$age < 11 , 2,
                              ifelse(hse08.mk2$age > 1 & hse08.mk2$age < 5 , 1, 0
)))
hse08.mk2$porftvg <- ifelse(hse08.mk2$porftvg > 5, 6, hse08.mk2$porftvg)
hse08.2 <- -c(which(hse08.mk2$aggr == 0))
hse08.mk2 <- hse08.mk2[hse08.2,]
hse08.mk2$age <- NULL

hse09.pc <- c("syslom","diaslom","sex","tenureb","age","origin","hhsz","addnum","IMD2007",
             "birthwt","porftvg","GOR07","sys2om","sys3om","dias2om","dias3om")
hse09.mk2 <- hse09[,hse09.pc]
colnames(hse09.mk2)[9] <- "imd"
colnames(hse09.mk2)[12] <- "gor"
hse09.mk2$year <- 09
hse09.mk2$aggr <- ifelse(hse09.mk2$age > 10 & hse09.mk2$age < 16 , 3,
                        ifelse(hse09.mk2$age > 4 & hse09.mk2$age < 11 , 2,
                              ifelse(hse09.mk2$age > 1 & hse09.mk2$age < 5 , 1, 0
)))
hse09.mk2$porftvg <- ifelse(hse09.mk2$porftvg > 5, 6, hse09.mk2$porftvg)
hse09.2 <- -c(which(hse09.mk2$aggr == 0))
hse09.mk2 <- hse09.mk2[hse09.2,]
hse09.mk2$age <- NULL

hse10.pc <- c("syslom","diaslom","sex","tenureb","age","origin","hhsz","addnum","imd2007",
             "birthwt","porftvg","gor1","sys2om","sys3om","dias2om","dias3om")
hse10.mk2 <- hse10[,hse10.pc]
colnames(hse10.mk2)[9] <- "imd"
colnames(hse10.mk2)[12] <- "gor"
hse10.mk2$year <- 10
hse10.mk2$aggr <- ifelse(hse10.mk2$age > 10 & hse10.mk2$age < 16 , 3,
                        ifelse(hse10.mk2$age > 4 & hse10.mk2$age < 11 , 2,
                              ifelse(hse10.mk2$age > 1 & hse10.mk2$age < 5 , 1, 0
)))
hse10.mk2$porftvg <- ifelse(hse10.mk2$porftvg > 5, 6, hse10.mk2$porftvg)
hse10.2 <- -c(which(hse10.mk2$aggr == 0))
hse10.mk2 <- hse10.mk2[hse10.2,]
hse10.mk2$age <- NULL

hse11.pc <- c("syslom","diaslom","Sex","tenureb","Age","Origin","HHSz","addnum","qim",
             "BirthWt","porftvg","gor1","sys2om","sys3om","dias2om","dias3om")

```



```

hse11.mk2 <- hse11[,hse11.pc]
colnames(hse11.mk2)[3] <- "sex"
colnames(hse11.mk2)[6] <- "origin"
colnames(hse11.mk2)[7] <- "hhszize"
colnames(hse11.mk2)[9] <- "imd"
colnames(hse11.mk2)[10] <- "birthwt"
colnames(hse11.mk2)[12] <- "gor"
hse11.mk2$year <- 11
hse11.mk2$aggr <- ifelse(hse11.mk2$Age > 10 & hse11.mk2$Age < 16 , 3,
                        ifelse(hse11.mk2$Age > 4 & hse11.mk2$Age < 11 , 2,
                              ifelse(hse11.mk2$Age > 1 & hse11.mk2$Age < 5 , 1, 0
)))
hse11.mk2$porftvg <- ifelse(hse11.mk2$porftvg > 5, 6, hse11.mk2$porftvg)
hse11.2 <- -c(which(hse11.mk2$aggr == 0))
hse11.mk2 <- hse11.mk2[hse11.2,]
hse11.mk2$Age <- NULL

hse12.pc <- c("syslom","diaslom","Sex","tenureb","Age","Origin","HHSzize","Addnum","qi
md",
             "BirthWt",          "gor1","sys2om","sys3om","dias2om","dias3om")
hse12.mk2 <- hse12[,hse12.pc]
colnames(hse12.mk2)[3] <- "sex"
colnames(hse12.mk2)[6] <- "origin"
colnames(hse12.mk2)[7] <- "hhszize"
colnames(hse12.mk2)[8] <- "addnum"
colnames(hse12.mk2)[9] <- "imd"
colnames(hse12.mk2)[10] <- "birthwt"
colnames(hse12.mk2)[11] <- "gor"
hse12.mk2$porftvg <- NA
hse12.mk2$year <- 12
hse12.mk2$aggr <- ifelse(hse12.mk2$Age > 10 & hse12.mk2$Age < 16 , 3,
                        ifelse(hse12.mk2$Age > 4 & hse12.mk2$Age < 11 , 2,
                              ifelse(hse12.mk2$Age > 1 & hse12.mk2$Age < 5 , 1, 0
)))
hse12.mk2$porftvg <- ifelse(hse12.mk2$porftvg > 5, 6, hse12.mk2$porftvg)
hse12.2 <- -c(which(hse12.mk2$aggr == 0))
hse12.mk2 <- hse12.mk2[hse12.2,]
hse12.mk2$Age <- NULL

hse13.pc <- c("SYS1OM","DIAS1OM","Sex","tenureb","Age","Origin","HHSzize","Addnum","qi
md",
             "BirthWt","porftvg","gor1","SYS2OM","SYS3OM","DIAS2OM","DIAS3OM")
hse13.mk2 <- hse13[,hse13.pc]
colnames(hse13.mk2)[1] <- "syslom"
colnames(hse13.mk2)[2] <- "diaslom"
colnames(hse13.mk2)[3] <- "sex"
colnames(hse13.mk2)[6] <- "origin"
colnames(hse13.mk2)[7] <- "hhszize"
colnames(hse13.mk2)[8] <- "addnum"
colnames(hse13.mk2)[9] <- "imd"
colnames(hse13.mk2)[10] <- "birthwt"
colnames(hse13.mk2)[12] <- "gor"
colnames(hse13.mk2)[13] <- "sys2om"
colnames(hse13.mk2)[14] <- "sys3om"
colnames(hse13.mk2)[15] <- "dias2om"
colnames(hse13.mk2)[16] <- "dias3om"

```

```

hse13.mk2$year <- 13
hse13.mk2$aggr <- ifelse(hse13.mk2$Age > 10 & hse13.mk2$Age < 16 , 3,
                        ifelse(hse13.mk2$Age > 4 & hse13.mk2$Age < 11 , 2,
                        ifelse(hse13.mk2$Age > 1 & hse13.mk2$Age < 5 , 1, 0
                        )))
hse13.mk2$porftvg <- ifelse(hse13.mk2$porftvg > 5, 6, hse13.mk2$porftvg)
hse13.2 <- -c(which(hse13.mk2$aggr == 0))
hse13.mk2 <- hse13.mk2[hse13.2,]
hse13.mk2$Age <- NULL

hse14.pc <- c("SYS10M","DIAS10M","Sex","tenureb","Age90","origin2","HHSIZE9","Addnum",
             "qimd",
             "BirthWt","PorFV05","gor1","SYS2OM","SYS3OM","DIAS2OM","DIAS3OM")
hse14.mk2 <- hse14[,hse14.pc]
colnames(hse14.mk2)[1] <- "sys1om"
colnames(hse14.mk2)[2] <- "dias1om"
colnames(hse14.mk2)[3] <- "sex"
colnames(hse14.mk2)[6] <- "origin"
colnames(hse14.mk2)[7] <- "hhsiz"
colnames(hse14.mk2)[8] <- "addnum"
colnames(hse14.mk2)[9] <- "imd"
colnames(hse14.mk2)[10] <- "birthwt"
colnames(hse14.mk2)[11] <- "porftvg"
colnames(hse14.mk2)[12] <- "gor"
colnames(hse14.mk2)[13] <- "sys2om"
colnames(hse14.mk2)[14] <- "sys3om"
colnames(hse14.mk2)[15] <- "dias2om"
colnames(hse14.mk2)[16] <- "dias3om"
hse14.mk2$year <- 14
hse14.mk2$aggr <- ifelse(hse14.mk2$Age90 > 10 & hse14.mk2$Age90 < 16 , 3,
                        ifelse(hse14.mk2$Age90 > 4 & hse14.mk2$Age90 < 11 , 2,
                        ifelse(hse14.mk2$Age90 > 1 & hse14.mk2$Age90 < 5 , 1,
                        0)))
hse14.2 <- -c(which(hse14.mk2$aggr == 0))
hse14.mk2 <- hse14.mk2[hse14.2,]
hse14.mk2$Age90 <- NULL

hse15.pc <- c("SYS10M","DIAS10M","Sex","tenureb","Ag015g4","origin2","HHSIZE6","addnu",
             "m","qimd",
             "BirthWt","PorFV05b","Gor1","SYS2OM","SYS3OM","DIAS2OM","DIAS3OM")
hse15.mk2 <- hse15[,hse15.pc]
colnames(hse15.mk2)[1] <- "sys1om"
colnames(hse15.mk2)[2] <- "dias1om"
colnames(hse15.mk2)[3] <- "sex"
colnames(hse15.mk2)[6] <- "origin"
colnames(hse15.mk2)[7] <- "hhsiz"
colnames(hse15.mk2)[9] <- "imd"
colnames(hse15.mk2)[10] <- "birthwt"
colnames(hse15.mk2)[11] <- "porftvg"
colnames(hse15.mk2)[12] <- "gor"
colnames(hse15.mk2)[13] <- "sys2om"
colnames(hse15.mk2)[14] <- "sys3om"
colnames(hse15.mk2)[15] <- "dias2om"
colnames(hse15.mk2)[16] <- "dias3om"
hse15.mk2$year <- 15
hse15.mk2$aggr <- ifelse(hse15.mk2$Ag015g4 > 0, hse15.mk2$Ag015g4, 0)

```

```
hse15.2 <- -c(which(hse15.mk2$aggr == 0))
hse15.mk2 <- hse15.mk2[hse15.2,]
hse15.mk2$Ag015g4 <- NULL
```

```
##### Data Preparation -- Merge (Row bind) the datasets #####
```

```
hse.mk20 <- rbind(hse05.mk2, hse06.mk2)
hse.mk20 <- rbind(hse.mk20, hse07.mk2)
hse.mk20 <- rbind(hse.mk20, hse08.mk2)
hse.mk20 <- rbind(hse.mk20, hse09.mk2)
hse.mk20 <- rbind(hse.mk20, hse10.mk2)
hse.mk20 <- rbind(hse.mk20, hse11.mk2)
hse.mk20 <- rbind(hse.mk20, hse12.mk2)
hse.mk20 <- rbind(hse.mk20, hse13.mk2)
hse.mk20 <- rbind(hse.mk20, hse14.mk2)
hse.mk20 <- rbind(hse.mk20, hse15.mk2)
```

```
##### Data Preparation -- Convert the missing values to NAs & Keep only the valid BP measurements #####
```

```
# As some of the missing observations are recorded as negative number or huge number in the datasets
# This step ensures those cases are marked as NAs rather than continuous number

colnames(hse.mk20)
```

```
## [1] "syslom" "diaslom" "sex" "tenureb" "origin" "hhsiz" "addnum"
## [8] "imd" "birthwt" "porftvg" "gor" "sys2om" "sys3om" "dias2om"
## [15] "dias3om" "year" "aggr"
```

```
range(hse.mk20$sex)
```

```
## [1] 1 2
```

```
range(hse.mk20$tenureb)
```

```
## [1] -9 6
```

```
range(hse.mk20$origin)
```

```
## [1] -9 18
```

```
range(hse.mk20$hhsiz)
```

```
## [1] 2 12
```

```
range(hse.mk20$addnum)
```

```
## [1] 1 56
```

```
range(hse.mk20$imd)
```

```
## [1] 1 5
```

```
range(hse.mk20$birthwt, na.rm = T)
```

```
## [1] -1.00 7.14
```

```
range(hse.mk20$porftvg, na.rm = T)
```

```
## [1] -9 6
```

```
range(hse.mk20$gor)
```

```
## [1] 1 9
```

```
range(hse.mk20$year)
```

```
## [1] 5 15
```

```
range(hse.mk20$aggr)
```

```
## [1] 1 3
```

```
range(hse.mk20$syslom)
```

```
## [1] -9 186
```

```
range(hse.mk20$sys2om)
```

```
## [1] -9 996
```

```
range(hse.mk20$sys3om)
```

```
## [1] -9 194
```

```
range(hse.mk20$diaslom)
```

```
## [1] -9 131
```

```
range(hse.mk20$dias2om)
```

```
## [1] -9 135
```

```
range(hse.mk20$dias3om)
```

```
## [1] -9 140
```

```
# tenureb, origin, birthwt, porftvg, & six BP measurements have negative/huge values  
# birthwt, porftvg have NA values originally which require to be imputed later.
```

```
#          Covert the missing values to NAs
```

```
hse.mk20$tenureb <- ifelse(hse.mk20$tenureb < 0, NA, hse.mk20$tenureb)  
hse.mk20$origin <- ifelse(hse.mk20$origin < 0, NA, hse.mk20$origin)  
hse.mk20$birthwt <- ifelse(hse.mk20$birthwt < 0, NA, hse.mk20$birthwt)  
hse.mk20$porftvg <- ifelse(hse.mk20$porftvg < 0, NA, hse.mk20$porftvg)
```

```
#          Keep only the valid measurements
```

```
#          Note: As children aged from 2-4 do not have valid BP measurements, so they are  
dropped from the analysis at this step
```

```
hse.mk20.pc <- -c(which(hse.mk20$syslom < 0 | hse.mk20$syslom > 200 |  
                      hse.mk20$sys2om < 0 | hse.mk20$sys2om > 200 |  
                      hse.mk20$sys3om < 0 | hse.mk20$sys3om > 200 |  
                      hse.mk20$diaslom < 0 | hse.mk20$diaslom > 200 |  
                      hse.mk20$dias2om < 0 | hse.mk20$dias2om > 200 |  
                      hse.mk20$dias3om < 0 | hse.mk20$dias3om > 200))  
hse.mk20 <- hse.mk20[hse.mk20.pc,]
```

```
#          Last check
```

```
colnames(hse.mk20)
```

```
## [1] "syslom" "diaslom" "sex"      "tenureb" "origin" "hhsizes" "addnum"  
## [8] "imd"    "birthwt" "porftvg" "gor"     "sys2om" "sys3om" "dias2om"  
## [15] "dias3om" "year"    "aggr"
```

```
range(hse.mk20$sex)
```

```
## [1] 1 2
```

```
range(hse.mk20$tenureb, na.rm = T)
```

```
## [1] 1 6
```

```
range(hse.mk20$origin, na.rm = T)
```

```
## [1] 1 18
```

```
range(hse.mk20$hhszsize)
```

```
## [1] 2 11
```

```
range(hse.mk20$addnum)
```

```
## [1] 1 56
```

```
range(hse.mk20$imd)
```

```
## [1] 1 5
```

```
range(hse.mk20$birthwt, na.rm = T)
```

```
## [1] 0.91 6.75
```

```
range(hse.mk20$porftvg, na.rm = T)
```

```
## [1] 0 6
```

```
range(hse.mk20$gor)
```

```
## [1] 1 9
```

```
range(hse.mk20$year)
```

```
## [1] 5 15
```

```
range(hse.mk20$aggr)
```

```
## [1] 2 3
```

```
range(hse.mk20$syslom, na.rm = T)
```

```
## [1] 51 183
```

```
range(hse.mk20$sys2om, na.rm = T)
```

```
## [1] 52 187
```

```
range(hse.mk20$sys3om, na.rm = T)
```

```
## [1] 53 194
```

```
range(hse.mk20$dias1om, na.rm = T)
```

```
## [1] 31 131
```

```
range(hse.mk20$dias2om, na.rm = T)
```

```
## [1] 31 135
```

```
range(hse.mk20$dias3om, na.rm = T)
```

```
## [1] 30 140
```

```
##### Data Preparation -- Average the BP for each row  
#####
```

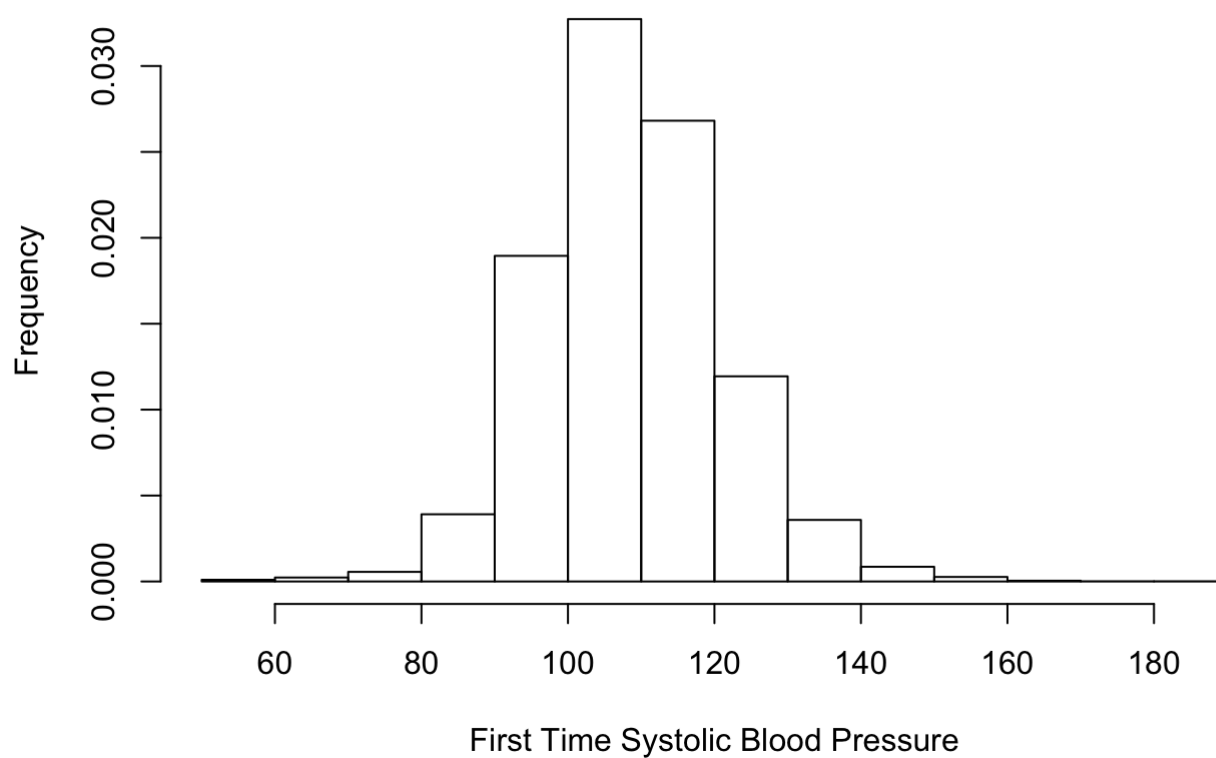
```
# Looking into the distribution of BP measurements
```

```
x1 <- hse.mk20$sys1om  
range(hse.mk20$sys1om, na.rm = T)
```

```
## [1] 51 183
```

```
hist(x1, freq = FALSE, main = "Distribution of First Time Systolic Blood Pressure Measurement",  
      , xlab = "First Time Systolic Blood Pressure",  
      , ylab = "Frequency")
```

Distribution of First Time Systolic Blood Pressure Measurement

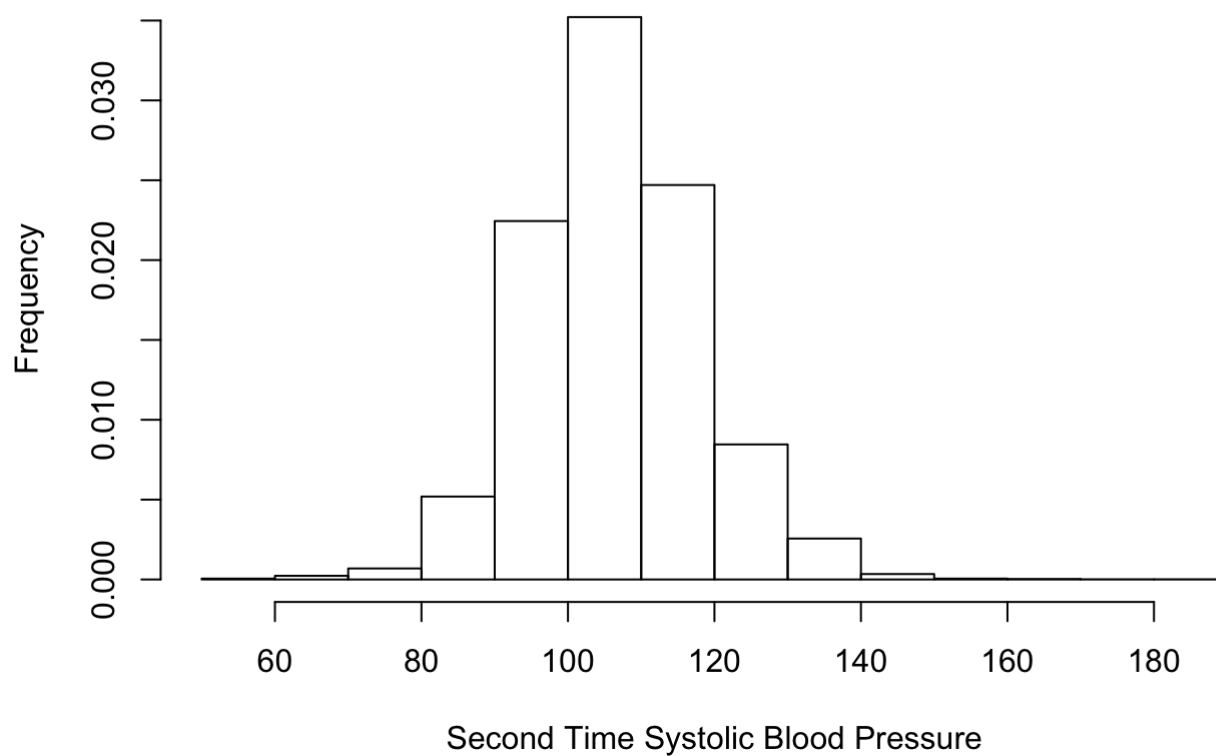


```
x2 <- hse.mk20$sys2om  
range(hse.mk20$sys2om, na.rm = T)
```

```
## [1] 52 187
```

```
hist(x2, freq = FALSE, main = "Distribution of Second Time Systolic Blood Pressure Me  
asurement"  
      , xlab = "Second Time Systolic Blood Pressure"  
      , ylab = "Frequency")
```


Distribution of Second Time Systolic Blood Pressure Measurement

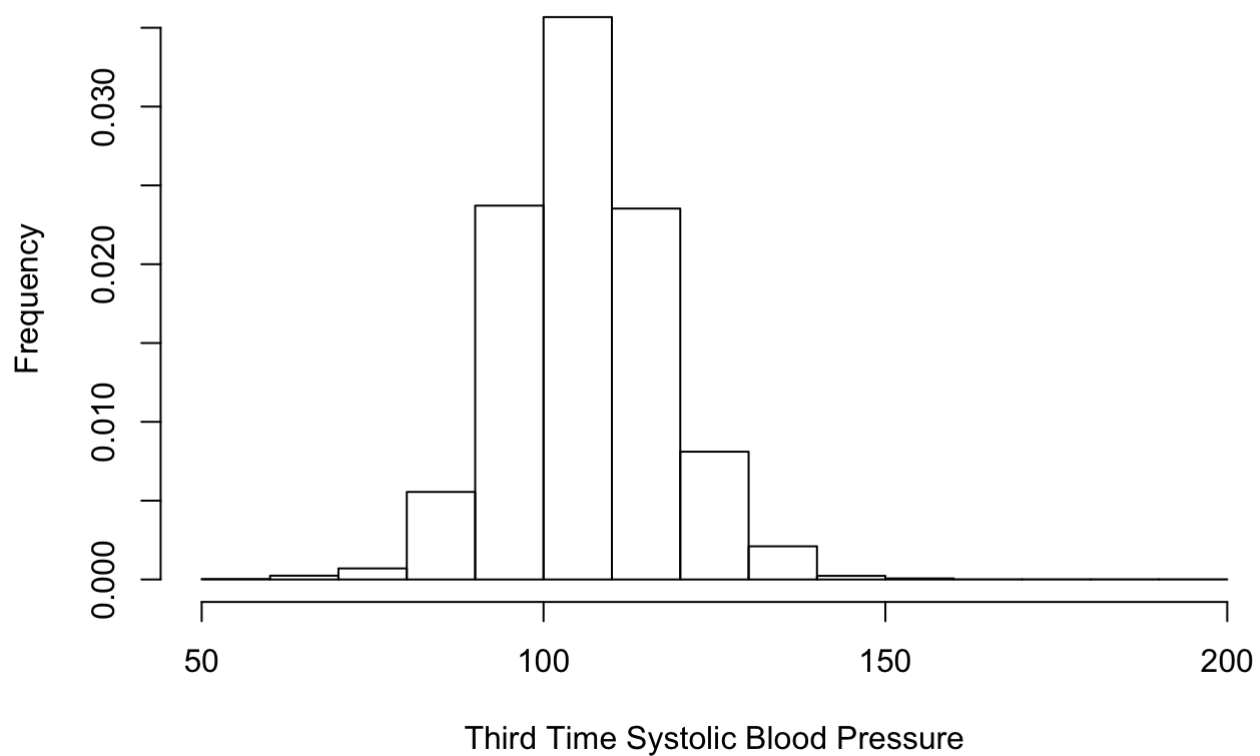


```
x3 <- hse.mk20$sys3om
range(hse.mk20$sys3om, na.rm = T)
```

```
## [1] 53 194
```

```
hist(x3, freq = FALSE, main = "Distribution of Third Time Systolic Blood Pressure Measurement",
      , xlab = "Third Time Systolic Blood Pressure",
      , ylab = "Frequency")
```

Distribution of Third Time Systolic Blood Pressure Measurement

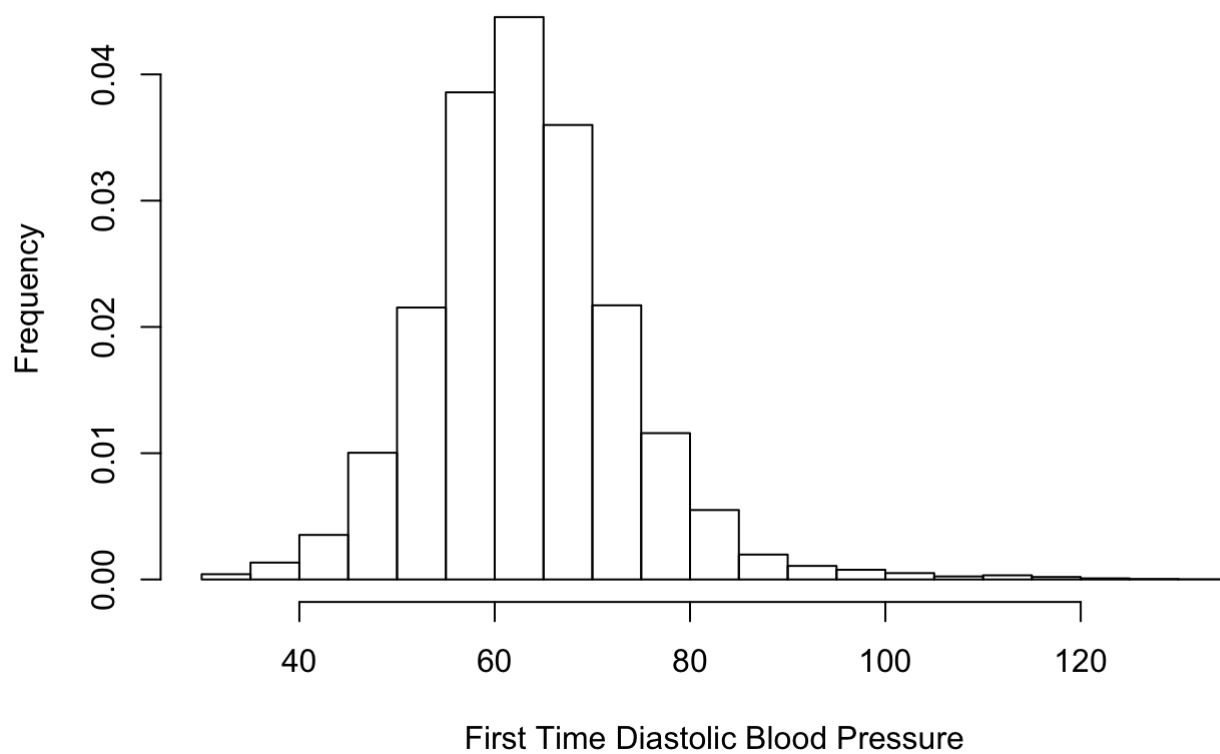


```
x4 <- hse.mk20$diaslom  
range(hse.mk20$diaslom, na.rm = T)
```

```
## [1] 31 131
```

```
hist(x4, freq = FALSE, main = "Distribution of First Time Diastolic Blood Pressure Me  
asurement"  
      , xlab = "First Time Diastolic Blood Pressure"  
      , ylab = "Frequency")
```

Distribution of First Time Diastolic Blood Pressure Measurement

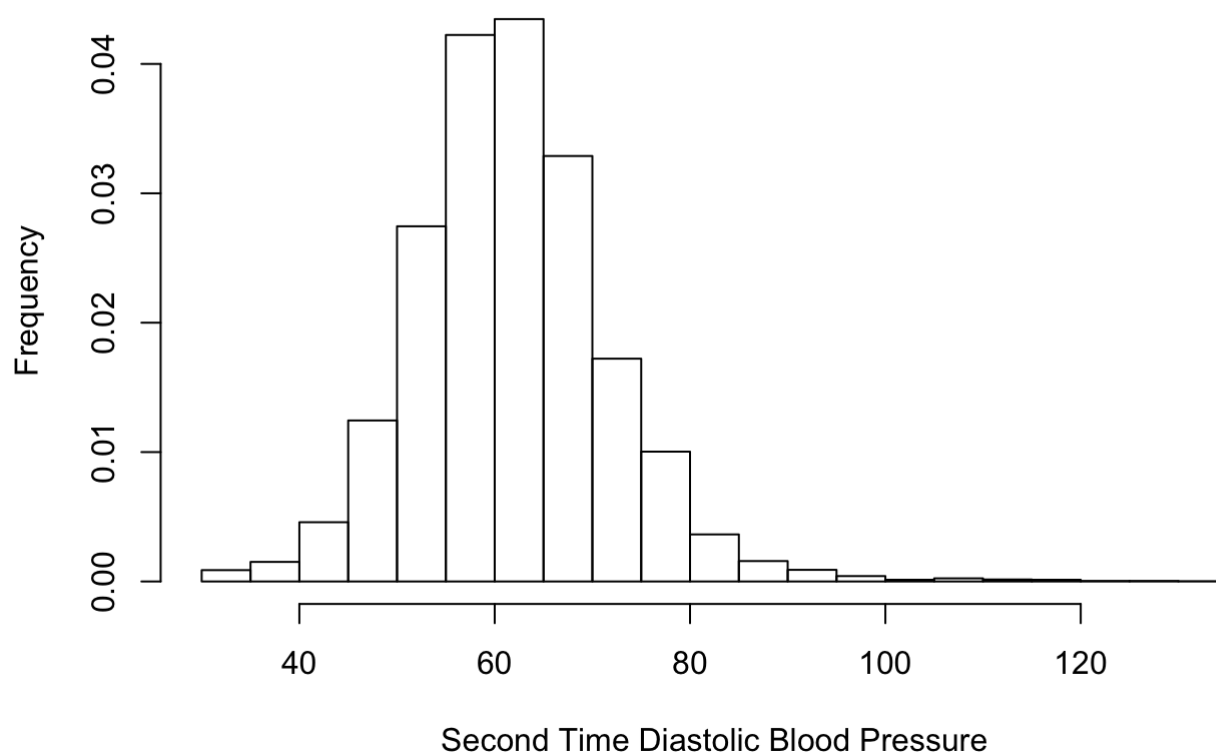


```
x5 <- hse.mk20$dias2om
range(hse.mk20$dias2om, na.rm = T)
```

```
## [1] 31 135
```

```
hist(x5, freq = FALSE, main = "Distribution of Second Time Diastolic Blood Pressure Measurement",
      , xlab = "Second Time Diastolic Blood Pressure",
      , ylab = "Frequency")
```

Distribution of Second Time Diastolic Blood Pressure Measurement

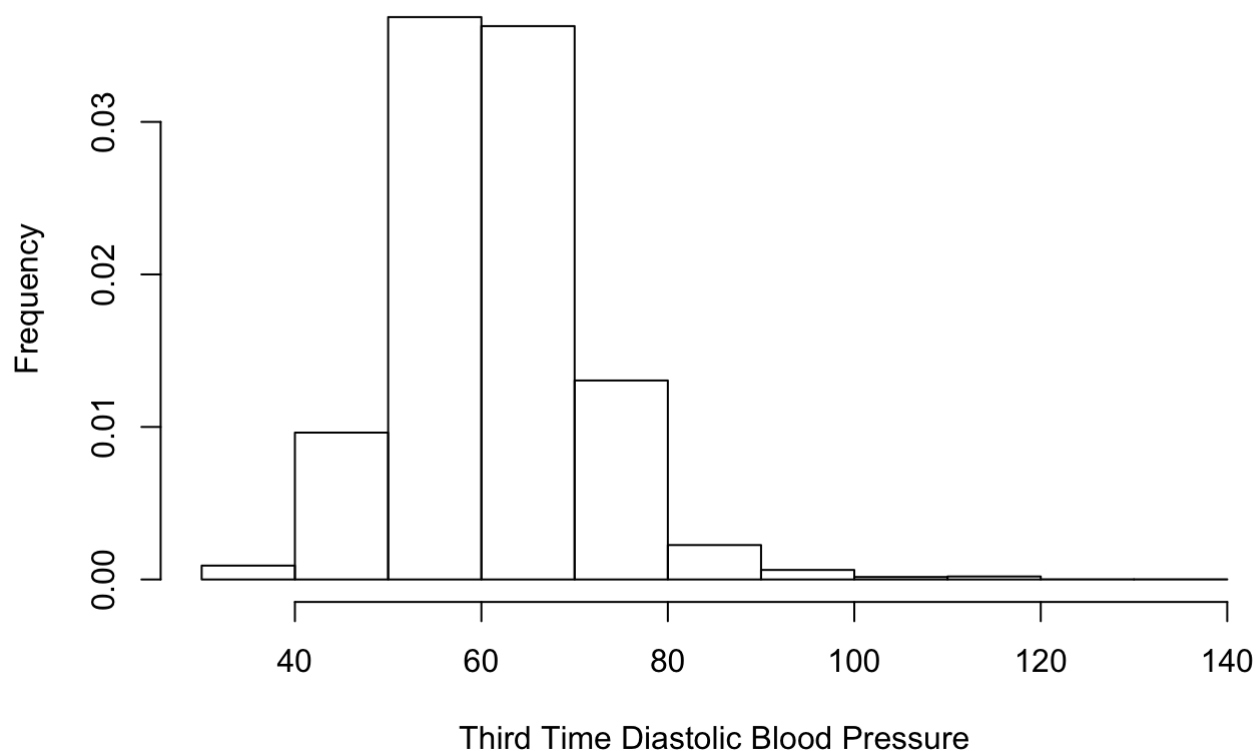


```
x6 <- hse.mk20$dias3om
range(hse.mk20$dias3om, na.rm = T)
```

```
## [1] 30 140
```

```
hist(x6, freq = FALSE, main = "Distribution of Third Time Diastolic Blood Pressure Measurement",
      , xlab = "Third Time Diastolic Blood Pressure",
      , ylab = "Frequency")
```

Distribution of Third Time Diastolic Blood Pressure Measurement



```
#      Get the average of the diastolic and systolic measurements for each row

hse.mk20$sysavg <- NA
hse.mk20$diaavg <- NA

for ( i in 1 : dim(hse.mk20)[1] ) {
  hse.mk20$sysavg[i] <- (hse.mk20$sys1om[i] + hse.mk20$sys2om[i] + hse.mk20$sys3om[i]) / 3
  hse.mk20$diaavg[i] <- (hse.mk20$dias1om[i] + hse.mk20$dias2om[i] + hse.mk20$dias3om[i]) / 3
}
hse.mk20$sys1om <- NULL
hse.mk20$sys2om <- NULL
hse.mk20$sys3om <- NULL
hse.mk20$dias1om <- NULL
hse.mk20$dias2om <- NULL
hse.mk20$dias3om <- NULL
```

```
##### Data Preparation -- Get the hypertensive group
#####

# Calculate .95 quantiles of the systolic and diastolic bp for each of age group,
# if the child's systolic / diastolic bp is higher than .95 age group's threshold,
# the child is classified as hypertensive

# systolic bp comparison for age group 2
hse.mk20$sys.hyper2 <- NA
hse.mk20$sys.hyper2 <- ifelse(hse.mk20$aggr == 2, hse.mk20$sysavg, NA)
range(hse.mk20$sys.hyper2, na.rm = T)
```

```
## [1] 66.33333 148.33333
```

```
quantile(hse.mk20$sys.hyper2, .95, na.rm = T)
```

```
##      95%
## 120.6667
```

```
hse.mk20$sys.hyper2.mk <- NA
hse.mk20$sys.hyper2.mk <- ifelse(hse.mk20$sys.hyper2 >= quantile(hse.mk20$sys.hyper2,
.95, na.rm = T)[1], 1, 0)
length(which(hse.mk20$sys.hyper2.mk == 1 ))
```

```
## [1] 240
```

```
# group 3
hse.mk20$sys.hyper3 <- NA
hse.mk20$sys.hyper3 <- ifelse(hse.mk20$aggr == 3, hse.mk20$sysavg, NA)
range(hse.mk20$sys.hyper3, na.rm = T)
```

```
## [1] 66 186
```

```
quantile(hse.mk20$sys.hyper3, .95, na.rm = T)
```

```
##      95%
## 129.6667
```

```
hse.mk20$sys.hyper3.mk <- NA
hse.mk20$sys.hyper3.mk <- ifelse(hse.mk20$sys.hyper3 >= quantile(hse.mk20$sys.hyper3,
.95, na.rm = T)[1], 1, 0)
length(which(hse.mk20$sys.hyper3.mk == 1 ))
```

```
## [1] 225
```

```
# diastolic bp comparison for age group 2
hse.mk20$dia.hyper2 <- NA
hse.mk20$dia.hyper2 <- ifelse(hse.mk20$aggr == 2, hse.mk20$diaavg, NA)
range(hse.mk20$dia.hyper2, na.rm = T)
```

```
## [1] 34.66667 102.00000
```

```
quantile(hse.mk20$dia.hyper2, .95, na.rm = T)
```

```
## 95%
## 77.33333
```

```
hse.mk20$dia.hyper2.mk <- NA
hse.mk20$dia.hyper2.mk <- ifelse(hse.mk20$dia.hyper2 >= quantile(hse.mk20$dia.hyper2,
.95, na.rm = T)[1], 1, 0)
length(which(hse.mk20$dia.hyper2.mk == 1 ))
```

```
## [1] 247
```

```
# group 3
hse.mk20$dia.hyper3 <- NA
hse.mk20$dia.hyper3 <- ifelse(hse.mk20$aggr == 3, hse.mk20$diaavg, NA)
range(hse.mk20$dia.hyper3, na.rm = T)
```

```
## [1] 35.66667 114.33333
```

```
quantile(hse.mk20$dia.hyper3, .95, na.rm = T)
```

```
## 95%
## 77
```

```
hse.mk20$dia.hyper3.mk <- NA
hse.mk20$dia.hyper3.mk <- ifelse(hse.mk20$dia.hyper3 >= quantile(hse.mk20$dia.hyper3,
.95, na.rm = T)[1], 1, 0)
length(which(hse.mk20$dia.hyper3.mk == 1 ))
```

```
## [1] 229
```

```

#           Get the hypertensive group
#   if the child is classified as hypertensive by one of the age groups'threshold
#   then the child is hypertensive

hse.mk20$sys.hyper2.mk <- ifelse(is.na(hse.mk20$sys.hyper2.mk), 0, hse.mk20$sys.hyper
2.mk)
hse.mk20$sys.hyper3.mk <- ifelse(is.na(hse.mk20$sys.hyper3.mk), 0, hse.mk20$sys.hyper
3.mk)

hse.mk20$dia.hyper2.mk <- ifelse(is.na(hse.mk20$dia.hyper2.mk), 0, hse.mk20$dia.hyper
2.mk)
hse.mk20$dia.hyper3.mk <- ifelse(is.na(hse.mk20$dia.hyper3.mk), 0, hse.mk20$dia.hyper
3.mk)

hse.mk20$hyper <- NA
hse.mk20$hyper <- ifelse(hse.mk20$sys.hyper2.mk == 1 | hse.mk20$sys.hyper3.mk == 1
                        | hse.mk20$dia.hyper2.mk == 1 | hse.mk20$dia.hyper3.mk == 1 ,1
, 0)
length(which(hse.mk20$hyper == 1 ))

```

```
## [1] 781
```

```

#           Delete used rows
hse.mk20$sys.hyper2 <- NULL
hse.mk20$sys.hyper2.mk <- NULL
hse.mk20$sys.hyper3 <- NULL
hse.mk20$sys.hyper3.mk <- NULL

hse.mk20$dia.hyper2 <- NULL
hse.mk20$dia.hyper2.mk <- NULL
hse.mk20$dia.hyper3 <- NULL
hse.mk20$dia.hyper3.mk <- NULL

```

```

##### Data Preparation -- Multiple Imputation & Comple
te cases #####
hse.mk50 <- hse.mk20

# Summary for which variables have missing data
summary(hse.mk50)

```



```
##          sex          tenureb          origin          hhsize
## Min.      :1.000    Min.      :1.000    Min.      : 1.000    Min.      : 2.000
## 1st Qu.:1.000    1st Qu.:2.000    1st Qu.: 1.000    1st Qu.: 3.000
## Median :1.000    Median :2.000    Median : 1.000    Median : 4.000
## Mean      :1.499    Mean      :2.573    Mean      : 1.979    Mean      : 4.087
## 3rd Qu.:2.000    3rd Qu.:4.000    3rd Qu.: 1.000    3rd Qu.: 5.000
## Max.      :2.000    Max.      :6.000    Max.      :18.000    Max.      :11.000
##
##          addnum          imd          birthwt          porftvg
## Min.      : 1.00    Min.      :1.000    Min.      :0.910    Min.      :0.000
## 1st Qu.: 7.00    1st Qu.:2.000    1st Qu.:2.980    1st Qu.:2.000
## Median :13.00    Median :3.000    Median :3.370    Median :3.000
## Mean      :16.27    Mean      :2.934    Mean      :3.343    Mean      :3.514
## 3rd Qu.:25.00    3rd Qu.:4.000    3rd Qu.:3.740    3rd Qu.:5.000
## Max.      :56.00    Max.      :5.000    Max.      :6.750    Max.      :6.000
##
##          gor          year          aggr          sysavg
## Min.      :1.000    Min.      : 5.0    Min.      :2.000    Min.      : 66.0
## 1st Qu.:3.000    1st Qu.: 6.0    1st Qu.:2.000    1st Qu.:100.3
## Median :5.000    Median : 8.0    Median :2.000    Median :107.0
## Mean      :5.195    Mean      : 9.5    Mean      :2.482    Mean      :107.6
## 3rd Qu.:8.000    3rd Qu.:13.0    3rd Qu.:3.000    3rd Qu.:114.0
## Max.      :9.000    Max.      :15.0    Max.      :3.000    Max.      :186.0
##
##          diaavg          hyper
## Min.      : 34.67    Min.      :0.00000
## 1st Qu.: 57.00    1st Qu.:0.00000
## Median : 62.33    Median :0.00000
## Mean      : 62.67    Mean      :0.08555
## 3rd Qu.: 68.00    3rd Qu.:0.00000
## Max.      :114.33    Max.      :1.00000
##
```

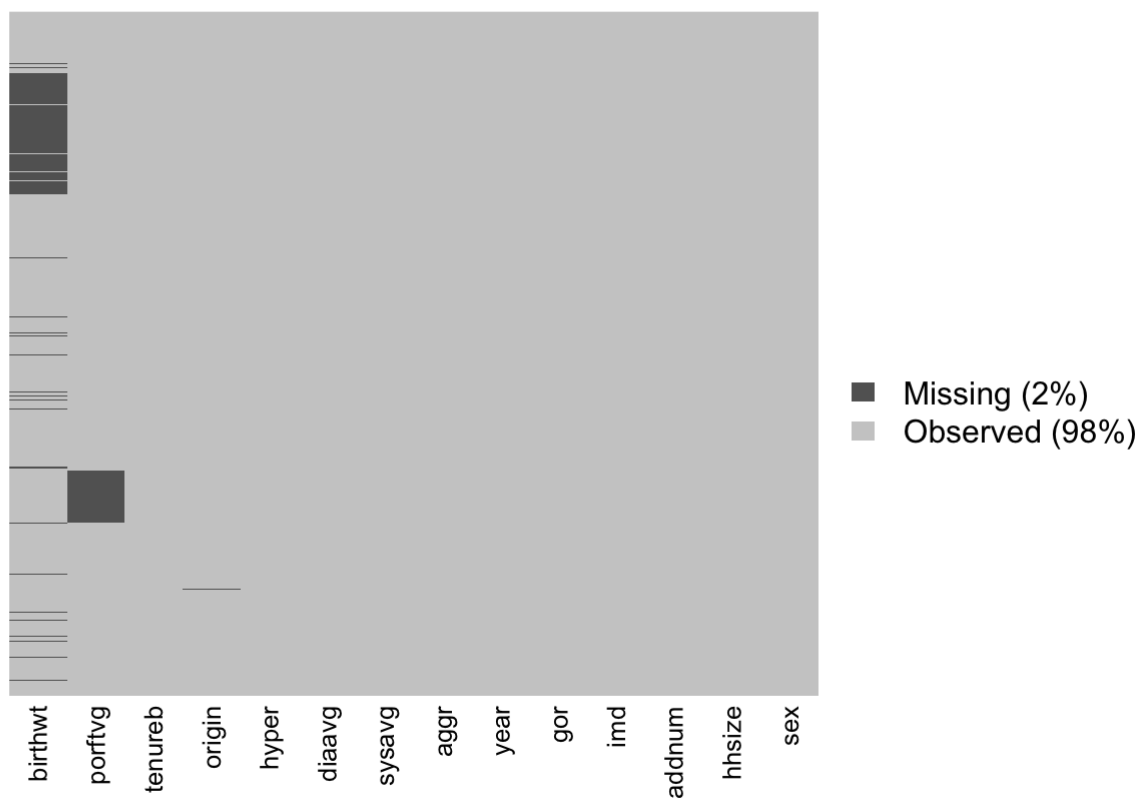
```
# Missing map to test if the missing data is missing at random
missmap(hse.mk50, col = c("grey39", "grey80"), y.labels = NULL, y.at = NULL)
```

```
## Warning in if (class(obj) == "amelia") {: the condition has length > 1 and
## only the first element will be used
```

```
## Warning: Unknown or uninitialised column: 'arguments'.
```

```
## Warning: Unknown or uninitialised column: 'arguments'.
```

Missingness Map



```
# tenure, origin and hyper can be considered to be missing at random, so they will not be
# imputed later, which will be removed from the analysis
# birth-weight and portvg cannot be considered to be missing at random, so they will
# be imputed

# Multiple imputation
set.seed(1)
hse.mk50.mi <- aregImpute(~ hyper + sex + hhsz + imd + birthwt +
                          portvg + aggr + year + sysavg + diaavg, data = hse.mk50,
                          n.impute = 50, nk=0)
```

```
## Iteration 1
Iteration 2
Iteration 3
Iteration 4
Iteration 5
Iteration 6
Iteration 7
Iteration 8
Iteration 9
Iteration 10
Iteration 11
Iteration 12
Iteration 13
Iteration 14
Iteration 15
Iteration 16
Iteration 17
Iteration 18
Iteration 19
Iteration 20
Iteration 21
Iteration 22
Iteration 23
Iteration 24
Iteration 25
Iteration 26
Iteration 27
Iteration 28
Iteration 29
Iteration 30
Iteration 31
Iteration 32
Iteration 33
Iteration 34
Iteration 35
Iteration 36
Iteration 37
Iteration 38
Iteration 39
Iteration 40
Iteration 41
Iteration 42
Iteration 43
Iteration 44
Iteration 45
Iteration 46
Iteration 47
Iteration 48
Iteration 49
Iteration 50
Iteration 51
Iteration 52
Iteration 53
```

```
# Retrieve the imputed values for birth weight
hse.mk50.mi.r <- impute.transcan(hse.mk50.mi, data = hse.mk50, imputation=1, list.out
=TRUE, pr=FALSE, check=FALSE)

# Arrange the columns accordingly
hse.mk60 <- hse.mk50
hse.mk60$birthwt <- hse.mk50.mi.r$birthwt
hse.mk60$porftvg <- hse.mk50.mi.r$porftvg

# Remove missing value in tenure, origin and porfvvg and( as they are missing at rando
m.
hse.mk60 <- hse.mk60[complete.cases(hse.mk60), ]
```

Part 2/3: The Effect of Deprivation – Propensity Analysis

```
##### Propensity Analysis #####
#####

# Covert the ordinal scale of treatment variable (imd) to binary scale
hse.mk90 <- hse.mk60

hse.mk90$imd <- ifelse(hse.mk90$imd == 1, 0,
                      ifelse(hse.mk90$imd == 2, 0,
                              ifelse(hse.mk90$imd == 3, NA,
                                      ifelse(hse.mk90$imd == 4, 1,
                                              ifelse(hse.mk90$imd == 5, 1, NA))))))

table(hse.mk90$imd)
```

```
##
##      0      1
## 3831 3492
```

```
hse.mk90 <- hse.mk90[complete.cases(hse.mk90), ]

# Compute the Propensity scores
reg <- glm(imd ~ sex + tenureb + origin + hhsize + addnum + birthwt + porftvg + gor +
aggr + year + hyper + sysavg + diaavg,
          family=binomial, data=hse.mk90)

hse.mk90$fit.value <- fitted.values(reg)

# Propensity Scores Matching & Average Treatment Effect on Treated
set.seed(8)
matching.vars <- cbind(hse.mk90$fit.value)
psm <- Match(Y=hse.mk90$hyper, Tr=hse.mk90$imd, X=matching.vars, Weight = 2, replace=
T, ties = F)
summary.Match(psm)
```

```
##  
## Estimate... 0.0189  
## SE..... 0.0068893  
## T-stat..... 2.7435  
## p.val..... 0.0060797  
##  
## Original number of observations..... 7323  
## Original number of treated obs..... 3492  
## Matched number of observations..... 3492  
## Matched number of observations (unweighted). 3492
```

```
# Balance test  
MatchBalance(imd ~ sex + tenureb + origin + hhsize + addnum + birthwt + porftvg + gor  
+ aggr + year,  
             match.out=psm, data=hse.mk90)
```

```

##
## ***** (V1) sex *****
##                               Before Matching      After Matching
## mean treatment.....         1.5032              1.5032
## mean control.....           1.4562              1.4876
## std mean diff.....          9.3917              3.1094
##
## mean raw eQQ diff.....       0.046964            0.01575
## med  raw eQQ diff.....        0                  0
## max  raw eQQ diff.....        1                  1
##
## mean eCDF diff.....          0.023482            0.0077745
## med  eCDF diff.....          0.023482            0.0077745
## max  eCDF diff.....          0.046964            0.015549
##
## var ratio (Tr/Co).....        1.0077              1.0006
## T-test p-value.....          7.9874e-05           0.18383
##
##
## ***** (V2) tenureb *****
##                               Before Matching      After Matching
## mean treatment.....          3.0003              3.0003
## mean control.....            2.2078              3.012
## std mean diff.....           72.022             -1.067
##
## mean raw eQQ diff.....       0.79496            0.11426
## med  raw eQQ diff.....        0                  0
## max  raw eQQ diff.....        2                  2
##
## mean eCDF diff.....          0.13245            0.019044
## med  eCDF diff.....          0.025508           0.025487
## max  eCDF diff.....          0.37388            0.034364
##
## var ratio (Tr/Co).....        1.6383              0.98222
## T-test p-value.....          < 2.22e-16           0.46043
## KS Bootstrap p-value.. < 2.22e-16              0.002
## KS Naive p-value..... < 2.22e-16           0.032369
## KS Statistic.....           0.37388            0.034364
##
##
## ***** (V3) origin *****
##                               Before Matching      After Matching
## mean treatment.....          2.5985              2.5985
## mean control.....            1.5583              2.5997
## std mean diff.....           29.889             -0.032915
##
## mean raw eQQ diff.....       1.0421              0.20619
## med  raw eQQ diff.....        0                  0
## max  raw eQQ diff.....        8                  3
##
## mean eCDF diff.....          0.057787            0.011455
## med  eCDF diff.....          0.055055           0.0094502
## max  eCDF diff.....          0.17262            0.049542
##
## var ratio (Tr/Co).....        2.9308              0.90676
## T-test p-value.....          < 2.22e-16           0.98706
## KS Bootstrap p-value.. < 2.22e-16           < 2.22e-16

```

```

## KS Naive p-value..... < 2.22e-16          0.00037915
## KS Statistic.....          0.17262          0.049542
##
##
## ***** (V4) hhsize *****
##                               Before Matching      After Matching
## mean treatment.....          4.1163          4.1163
## mean control.....          4.0733          4.0819
## std mean diff.....          3.409          2.7297
##
## mean raw eQQ diff.....      0.31501          0.22279
## med  raw eQQ diff.....          0          0
## max  raw eQQ diff.....          2          1
##
## mean eCDF diff.....          0.031467          0.022279
## med  eCDF diff.....          0.019711          0.0098797
## max  eCDF diff.....          0.087419          0.067583
##
## var ratio (Tr/Co).....          1.8663          1.4272
## T-test p-value.....          0.098741          0.21158
## KS Bootstrap p-value.. < 2.22e-16          < 2.22e-16
## KS Naive p-value.....      1.4954e-12          2.367e-07
## KS Statistic.....          0.087419          0.067583
##
##
## ***** (V5) addnum *****
##                               Before Matching      After Matching
## mean treatment.....          16.173          16.173
## mean control.....          16.14          15.542
## std mean diff.....          0.2755          5.3417
##
## mean raw eQQ diff.....      0.23883          0.72766
## med  raw eQQ diff.....          0          1
## max  raw eQQ diff.....          7          7
##
## mean eCDF diff.....          0.0046516          0.015989
## med  eCDF diff.....          0.0036061          0.015607
## max  eCDF diff.....          0.014928          0.046964
##
## var ratio (Tr/Co).....          0.97464          1.0161
## T-test p-value.....          0.90684          0.023588
## KS Bootstrap p-value..          0.626          < 2.22e-16
## KS Naive p-value.....          0.81029          0.00090369
## KS Statistic.....          0.014928          0.046964
##
##
## ***** (V6) birthwt *****
##                               Before Matching      After Matching
## mean treatment.....          3.2719          3.2719
## mean control.....          3.4058          3.2783
## std mean diff.....          -21.015          -1.0015
##
## mean raw eQQ diff.....      0.13314          0.039462
## med  raw eQQ diff.....          0.12          0.03
## max  raw eQQ diff.....          0.61          0.56
##
## mean eCDF diff.....          0.040737          0.010641
## med  eCDF diff.....          0.035587          0.0080183

```

```

## max eCDF diff..... 0.098035 0.032932
##
## var ratio (Tr/Co)..... 1.0899 0.97599
## T-test p-value..... < 2.22e-16 0.67053
## KS Bootstrap p-value.. < 2.22e-16 0.02
## KS Naive p-value..... 1.1102e-15 0.045316
## KS Statistic..... 0.098035 0.032932
##
##
## ***** (V7) porftvg *****
## Before Matching After Matching
## mean treatment..... 3.3562 3.3562
## mean control..... 3.7001 3.3932
## std mean diff..... -19.935 -2.1419
##
## mean raw eQQ diff..... 0.34278 0.040951
## med raw eQQ diff..... 0 0
## max raw eQQ diff..... 2 1
##
## mean eCDF diff..... 0.049119 0.0058501
## med eCDF diff..... 0.043448 0.0037228
## max eCDF diff..... 0.078635 0.011168
##
## var ratio (Tr/Co)..... 1.0969 1.0123
## T-test p-value..... < 2.22e-16 0.3576
## KS Bootstrap p-value.. < 2.22e-16 0.714
## KS Naive p-value..... 3.0848e-10 0.98138
## KS Statistic..... 0.078635 0.011168
##
##
## ***** (V8) gor *****
## Before Matching After Matching
## mean treatment..... 4.5326 4.5326
## mean control..... 5.7022 4.4948
## std mean diff..... -45.873 1.4827
##
## mean raw eQQ diff..... 1.1687 0.17125
## med raw eQQ diff..... 1 0
## max raw eQQ diff..... 2 1
##
## mean eCDF diff..... 0.12995 0.019028
## med eCDF diff..... 0.15147 0.0088774
## max eCDF diff..... 0.20495 0.075029
##
## var ratio (Tr/Co)..... 1.0181 1.0054
## T-test p-value..... < 2.22e-16 0.4754
## KS Bootstrap p-value.. < 2.22e-16 < 2.22e-16
## KS Naive p-value..... < 2.22e-16 5.8061e-09
## KS Statistic..... 0.20495 0.075029
##
##
## ***** (V9) aggr *****
## Before Matching After Matching
## mean treatment..... 2.4653 2.4653
## mean control..... 2.4978 2.4633
## std mean diff..... -6.5011 0.40183
##
## mean raw eQQ diff..... 0.03236 0.0020046

```



```
## med   raw eQQ diff.....      0              0
## max   raw eQQ diff.....      1              1
##
## mean  eCDF diff.....      0.016216      0.0010023
## med   eCDF diff.....      0.016216      0.0010023
## max   eCDF diff.....      0.032432      0.0020046
##
## var ratio (Tr/Co).....      0.99524      1.0006
## T-test p-value.....      0.005526      0.86617
##
##
## ***** (V10) year *****
##                               Before Matching      After Matching
## mean treatment.....      9.5745      9.5745
## mean control.....      9.4576      9.6418
## std mean diff.....      3.5328      -2.0342
##
## mean raw eQQ diff.....      0.11856      0.2815
## med   raw eQQ diff.....      0              0
## max   raw eQQ diff.....      2              2
##
## mean eCDF diff.....      0.011536      0.025258
## med   eCDF diff.....      0.0093885      0.029353
## max   eCDF diff.....      0.02675      0.053551
##
## var ratio (Tr/Co).....      1.0424      1.1379
## T-test p-value.....      0.12735      0.37932
## KS Bootstrap p-value..      0.054      < 2.22e-16
## KS Naive p-value.....      0.14638      8.9535e-05
## KS Statistic.....      0.02675      0.053551
##
##
## Before Matching Minimum p.value: < 2.22e-16
## Variable Name(s): tenureb origin hhsize birthwt porftvg gor   Number(s): 2 3 4 6 7
8
##
## After Matching Minimum p.value: < 2.22e-16
## Variable Name(s): origin hhsize addnum gor year   Number(s): 3 4 5 8 10
```

```
# Create a new dataset after matching
hse.mk.FX <- rbind(hse.mk90[psm$index.control,],hse.mk90[psm$index.treated,])

# Regression with a matched dataset
ATT <- lm(hyper ~ imd, data=hse.mk.FX)
summary(ATT)
```

```
##
## Call:
## lm(formula = hyper ~ imd, data = hse.mk.FX)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.10281 -0.10281 -0.08391 -0.08391  0.91609
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.083906   0.004921  17.049 < 2e-16 ***
## imd          0.018900   0.006960   2.716  0.00663 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2908 on 6982 degrees of freedom
## Multiple R-squared:  0.001055,    Adjusted R-squared:  0.000912
## F-statistic: 7.375 on 1 and 6982 DF,  p-value: 0.006631
```

```
# Comprison with Linear Regression result from unmatched dataset
LR <- lm(hyper ~ imd, data=hse.mk90)
summary(LR)
```

```
##
## Call:
## lm(formula = hyper ~ imd, data = hse.mk90)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.10281 -0.10281 -0.07387 -0.07387  0.92613
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.073871   0.004564  16.186 < 2e-16 ***
## imd          0.028935   0.006609   4.378 1.21e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2825 on 7321 degrees of freedom
## Multiple R-squared:  0.002611,    Adjusted R-squared:  0.002475
## F-statistic: 19.17 on 1 and 7321 DF,  p-value: 1.214e-05
```

```
# Baseline patient characteristics BEFORE propensity score matching
attr(hse.mk90$birthwt, "class") <- c("numeric")

hse.mk90.bn <- hse.mk90[c(hse.mk90$imd == 0),]
myVars <- c("sex", "tenureb", "origin", "hhsz", "birthwt", "porftvg", "gor", "aggr",
"year")
catVars <- c("sex", "tenureb", "origin", "hhsz", "porftvg", "gor", "aggr", "year")
tab1 <- CreateTableOne(vars = myVars, data = hse.mk90.bn, factorVars = catVars)
tab1
```

```
##
##                                Overall
##      n                        3831
##      sex = 2 (%)              1868 (48.8)
##      tenureb (%)
##          1                    448 (11.7)
##          2                    2766 (72.2)
##          3                     15 ( 0.4)
##          4                    577 (15.1)
##          5                     25 ( 0.7)
##      origin (%)
##          1                    3404 (88.9)
##          2                     42 ( 1.1)
##          3                    109 ( 2.8)
##          4                     81 ( 2.1)
##          5                     19 ( 0.5)
##          6                     19 ( 0.5)
##          7                     24 ( 0.6)
##          8                     38 ( 1.0)
##          9                     27 ( 0.7)
##         10                     6 ( 0.2)
##         11                     5 ( 0.1)
##         12                     9 ( 0.2)
##         13                     16 ( 0.4)
##         14                     16 ( 0.4)
##         15                     9 ( 0.2)
##         16                     4 ( 0.1)
##         17                     1 ( 0.0)
##         18                     2 ( 0.1)
##      hhsize (%)
##          2                    138 ( 3.6)
##          3                    690 (18.0)
##          4                   2048 (53.5)
##          5                    732 (19.1)
##          6                    179 ( 4.7)
##          7                     24 ( 0.6)
##          8                     16 ( 0.4)
##          9                     3 ( 0.1)
##         10                     1 ( 0.0)
##      birthwt (mean (SD)) 3.41 (0.61)
##      porftvg (%)
##          0                    157 ( 4.1)
##          1                    124 ( 3.2)
##          2                    675 (17.6)
##          3                    852 (22.2)
##          4                    719 (18.8)
##          5                    555 (14.5)
##          6                    749 (19.6)
##      gor (%)
##          1                    184 ( 4.8)
##          2                    403 (10.5)
##          3                    351 ( 9.2)
##          4                    420 (11.0)
##          5                    293 ( 7.6)
##          6                    547 (14.3)
##          7                    262 ( 6.8)
##          8                    798 (20.8)
```

```
##          9                573 (15.0)
##   aggr = 3 (%)          1907 (49.8)
##   year (%)
##          5                361 ( 9.4)
##          6                661 (17.3)
##          7                317 ( 8.3)
##          8                613 (16.0)
##         10                337 ( 8.8)
##         11                325 ( 8.5)
##         12                297 ( 7.8)
##         13                330 ( 8.6)
##         14                311 ( 8.1)
##         15                279 ( 7.3)
```

```
hse.mk90.by <- hse.mk90[c(hse.mk90$imd == 1),]
myVars <- c("sex", "tenureb", "origin", "hhsz", "birthwt", "porftvg", "gor", "aggr",
"year")
catVars <- c("sex", "tenureb", "origin", "hhsz", "porftvg", "gor", "aggr", "year")
tab2 <- CreateTableOne(vars = myVars, data = hse.mk90.by, factorVars = catVars)
tab2
```

```

##
## Overall
## n 3492
## sex = 2 (%) 1757 (50.3)
## tenureb (%)
## 1 234 ( 6.7)
## 2 1390 (39.8)
## 3 29 ( 0.8)
## 4 1820 (52.1)
## 5 18 ( 0.5)
## 6 1 ( 0.0)
## origin (%)
## 1 2500 (71.6)
## 2 117 ( 3.4)
## 3 259 ( 7.4)
## 4 158 ( 4.5)
## 5 36 ( 1.0)
## 6 13 ( 0.4)
## 7 29 ( 0.8)
## 8 31 ( 0.9)
## 9 73 ( 2.1)
## 10 49 ( 1.4)
## 11 41 ( 1.2)
## 12 18 ( 0.5)
## 13 66 ( 1.9)
## 14 44 ( 1.3)
## 15 22 ( 0.6)
## 16 20 ( 0.6)
## 17 4 ( 0.1)
## 18 12 ( 0.3)
## hhsize (%)
## 2 295 ( 8.4)
## 3 765 (21.9)
## 4 1303 (37.3)
## 5 711 (20.4)
## 6 270 ( 7.7)
## 7 100 ( 2.9)
## 8 36 ( 1.0)
## 9 7 ( 0.2)
## 10 4 ( 0.1)
## 11 1 ( 0.0)
## birthwt (mean (SD)) 3.27 (0.64)
## porftvg (%)
## 0 269 ( 7.7)
## 1 138 ( 4.0)
## 2 739 (21.2)
## 3 774 (22.2)
## 4 609 (17.4)
## 5 432 (12.4)
## 6 531 (15.2)
## gor (%)
## 1 424 (12.1)
## 2 640 (18.3)
## 3 433 (12.4)
## 4 321 ( 9.2)
## 5 364 (10.4)
## 6 207 ( 5.9)

```

```
##          7          569 (16.3)
##          8          310 ( 8.9)
##          9          224 ( 6.4)
##   aggr = 3 (%)      1625 (46.5)
##   year (%)
##          5          305 ( 8.7)
##          6          612 (17.5)
##          7          262 ( 7.5)
##          8          595 (17.0)
##         10          293 ( 8.4)
##         11          232 ( 6.6)
##         12          261 ( 7.5)
##         13          334 ( 9.6)
##         14          283 ( 8.1)
##         15          315 ( 9.0)
```

```
# Baseline patient characteristics AFTER propensity score matching
```

```
attr(hse.mk.FX$birthwt, "class") <- c("numeric")
```

```
hse.mk.FX.an <- hse.mk.FX[c(hse.mk.FX$imd == 0),]
```

```
myVars <- c("sex", "tenureb", "origin", "hhsize", "birthwt", "porftvg", "gor", "aggr",
"year")
```

```
catVars <- c("sex", "tenureb", "origin", "hhsize", "porftvg", "gor", "aggr", "year")
```

```
tab3 <- CreateTableOne(vars = myVars, data = hse.mk.FX.an, factorVars = catVars)
```

```
tab3
```

```
##
##                                Overall
##      n                        3492
##      sex = 2 (%)              1593 (45.6)
##      tenureb (%)
##          1                    134 ( 3.8)
##          2                    1590 (45.5)
##          3                      7 ( 0.2)
##          4                    1622 (46.4)
##          5                     139 ( 4.0)
##      origin (%)
##          1                    2673 (76.5)
##          2                      62 ( 1.8)
##          3                     150 ( 4.3)
##          4                      90 ( 2.6)
##          5                      43 ( 1.2)
##          6                      31 ( 0.9)
##          7                      55 ( 1.6)
##          8                      78 ( 2.2)
##          9                      42 ( 1.2)
##         10                      7 ( 0.2)
##         11                      9 ( 0.3)
##         12                      16 ( 0.5)
##         13                     102 ( 2.9)
##         14                      70 ( 2.0)
##         15                      25 ( 0.7)
##         16                      29 ( 0.8)
##         17                      3 ( 0.1)
##         18                      7 ( 0.2)
##      hhsize (%)
##          2                     169 ( 4.8)
##          3                     688 (19.7)
##          4                    1742 (49.9)
##          5                     609 (17.4)
##          6                     197 ( 5.6)
##          7                      47 ( 1.3)
##          8                      34 ( 1.0)
##          9                      4 ( 0.1)
##         10                      2 ( 0.1)
##      birthwt (mean (SD)) 3.28 (0.64)
##      porftvg (%)
##          0                     230 ( 6.6)
##          1                     184 ( 5.3)
##          2                     697 (20.0)
##          3                     796 (22.8)
##          4                     583 (16.7)
##          5                     461 (13.2)
##          6                     541 (15.5)
##      gor (%)
##          1                     442 (12.7)
##          2                     609 (17.4)
##          3                     400 (11.5)
##          4                     452 (12.9)
##          5                     248 ( 7.1)
##          6                     500 (14.3)
##          7                     183 ( 5.2)
##          8                     415 (11.9)
```

```
##          9                243 ( 7.0)
##   aggr = 3 (%)          1618 (46.3)
##   year (%)
##          5                280 ( 8.0)
##          6                512 (14.7)
##          7                200 ( 5.7)
##          8                681 (19.5)
##         10                324 ( 9.3)
##         11                338 ( 9.7)
##         12                349 (10.0)
##         13                314 ( 9.0)
##         14                289 ( 8.3)
##         15                205 ( 5.9)
```

```
hse.mk.FX.ay <- hse.mk.FX[c(hse.mk.FX$imd == 1),]
myVars <- c("sex", "tenureb", "origin", "hhsz", "birthwt", "porftvg", "gor", "aggr",
"year")
catVars <- c("sex", "tenureb", "origin", "hhsz", "porftvg", "gor", "aggr", "year")
tab4 <- CreateTableOne(vars = myVars, data = hse.mk.FX.ay, factorVars = catVars)
tab4
```



```

##
## Overall
## n 3492
## sex = 2 (%) 1757 (50.3)
## tenureb (%)
## 1 234 ( 6.7)
## 2 1390 (39.8)
## 3 29 ( 0.8)
## 4 1820 (52.1)
## 5 18 ( 0.5)
## 6 1 ( 0.0)
## origin (%)
## 1 2500 (71.6)
## 2 117 ( 3.4)
## 3 259 ( 7.4)
## 4 158 ( 4.5)
## 5 36 ( 1.0)
## 6 13 ( 0.4)
## 7 29 ( 0.8)
## 8 31 ( 0.9)
## 9 73 ( 2.1)
## 10 49 ( 1.4)
## 11 41 ( 1.2)
## 12 18 ( 0.5)
## 13 66 ( 1.9)
## 14 44 ( 1.3)
## 15 22 ( 0.6)
## 16 20 ( 0.6)
## 17 4 ( 0.1)
## 18 12 ( 0.3)
## hhsize (%)
## 2 295 ( 8.4)
## 3 765 (21.9)
## 4 1303 (37.3)
## 5 711 (20.4)
## 6 270 ( 7.7)
## 7 100 ( 2.9)
## 8 36 ( 1.0)
## 9 7 ( 0.2)
## 10 4 ( 0.1)
## 11 1 ( 0.0)
## birthwt (mean (SD)) 3.27 (0.64)
## porftvg (%)
## 0 269 ( 7.7)
## 1 138 ( 4.0)
## 2 739 (21.2)
## 3 774 (22.2)
## 4 609 (17.4)
## 5 432 (12.4)
## 6 531 (15.2)
## gor (%)
## 1 424 (12.1)
## 2 640 (18.3)
## 3 433 (12.4)
## 4 321 ( 9.2)
## 5 364 (10.4)
## 6 207 ( 5.9)

```

```
##          7          569 (16.3)
##          8          310 ( 8.9)
##          9          224 ( 6.4)
##   aggr = 3 (%)      1625 (46.5)
##   year (%)
##          5          305 ( 8.7)
##          6          612 (17.5)
##          7          262 ( 7.5)
##          8          595 (17.0)
##         10          293 ( 8.4)
##         11          232 ( 6.6)
##         12          261 ( 7.5)
##         13          334 ( 9.6)
##         14          283 ( 8.1)
##         15          315 ( 9.0)
```

Part 3/3: Propensity Analysis – Sensitivity Test

```
##### Sensitivity Test #####
#####
```

```
# If one treated observation matches more than one control observation,
# the matched dataset will include the multiple matched control observations and
# the matched data will be weighted to reflect the multiple matches.
# The sum of the weighted observations will still equal the original number of observations.
```

```
# Propensity Scores Matching & Average Treatment Effect on Treated
psm <- Match(Y=hse.mk90$hyper, Tr=hse.mk90$imd, X=matching.vars, Weight = 2, replace=
T, ties = T, version = "fast")
summary.Match(psm)
```

```
##
## Estimate... 0.012838
## SE..... 0.0070846
## T-stat..... 1.8121
## p.val..... 0.069974
##
## Original number of observations..... 7323
## Original number of treated obs..... 3492
## Matched number of observations..... 3492
## Matched number of observations (unweighted). 20535
```

```
# Create a new dataset after matching
hse.mk.FX <- rbind(hse.mk90[psm$index.control,],hse.mk90[psm$index.treated,])

# Regression with a matched dataset
ATT.S <- lm(hyper ~ imd, data=hse.mk.FX)
summary(ATT.S)
```

```
##
## Call:
## lm(formula = hyper ~ imd, data = hse.mk.FX)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.08381 -0.08381 -0.08225 -0.08225  0.91775
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.082250   0.001926  42.715  <2e-16 ***
## imd          0.001558   0.002723   0.572    0.567
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2759 on 41068 degrees of freedom
## Multiple R-squared:  7.974e-06, Adjusted R-squared: -1.638e-05
## F-statistic: 0.3275 on 1 and 41068 DF, p-value: 0.5672
```