

BSc Population Health Dissertation (18/19)

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(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()  
## ✖ 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: lattice
```

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

```
## 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
```

```
##### 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", "hhsize", "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", "hhsize", "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", "hhsized", "addnum", "imd2007",
             "birthwt", "porftvg", "gor07", "sys2om", "sys3om", "dias2om", "dias3om")
hse07.mk2 <- hse07[,hse07.pc]
colnames(hse07.mk2)[6] <- "origin"
colnames(hse07.mk2)[7] <- "hhsize"
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] <- "hhszize"
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] <- "hhszize"
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$hhsz)
```

```
## [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  
#####
```

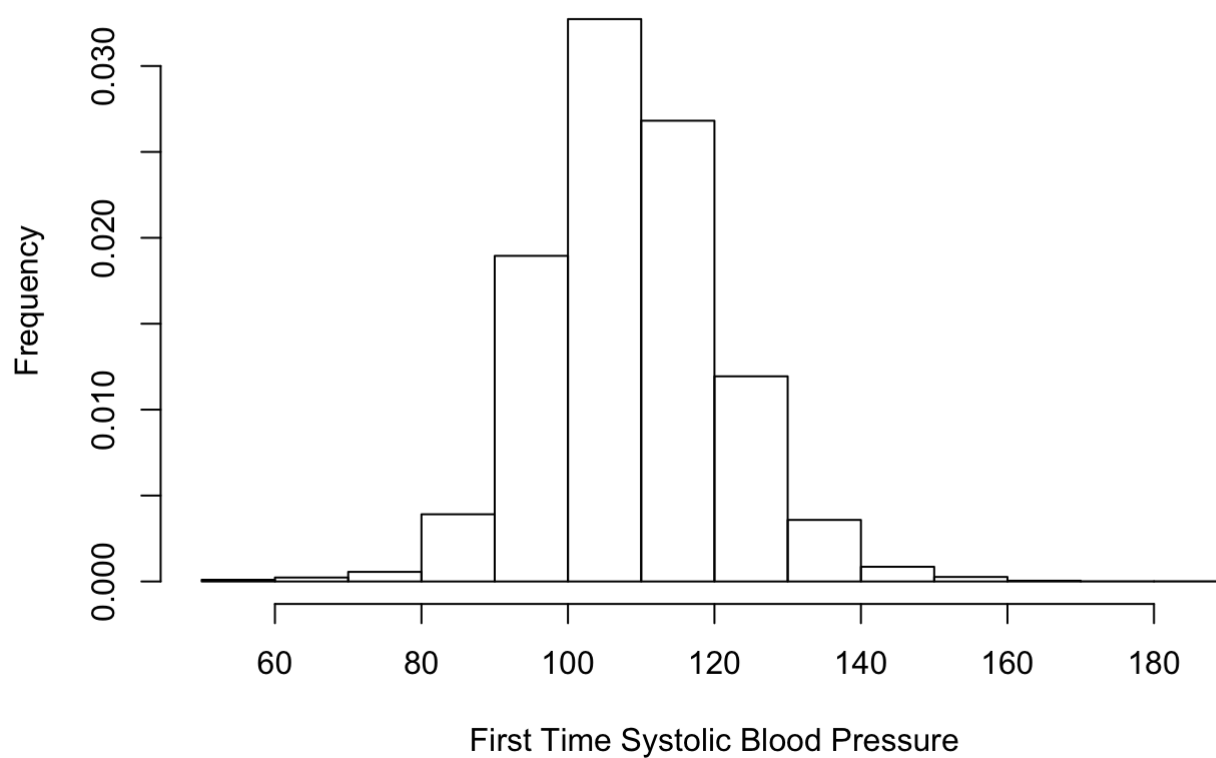
```
# Looking into the distribution of BP measurements
```

```
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

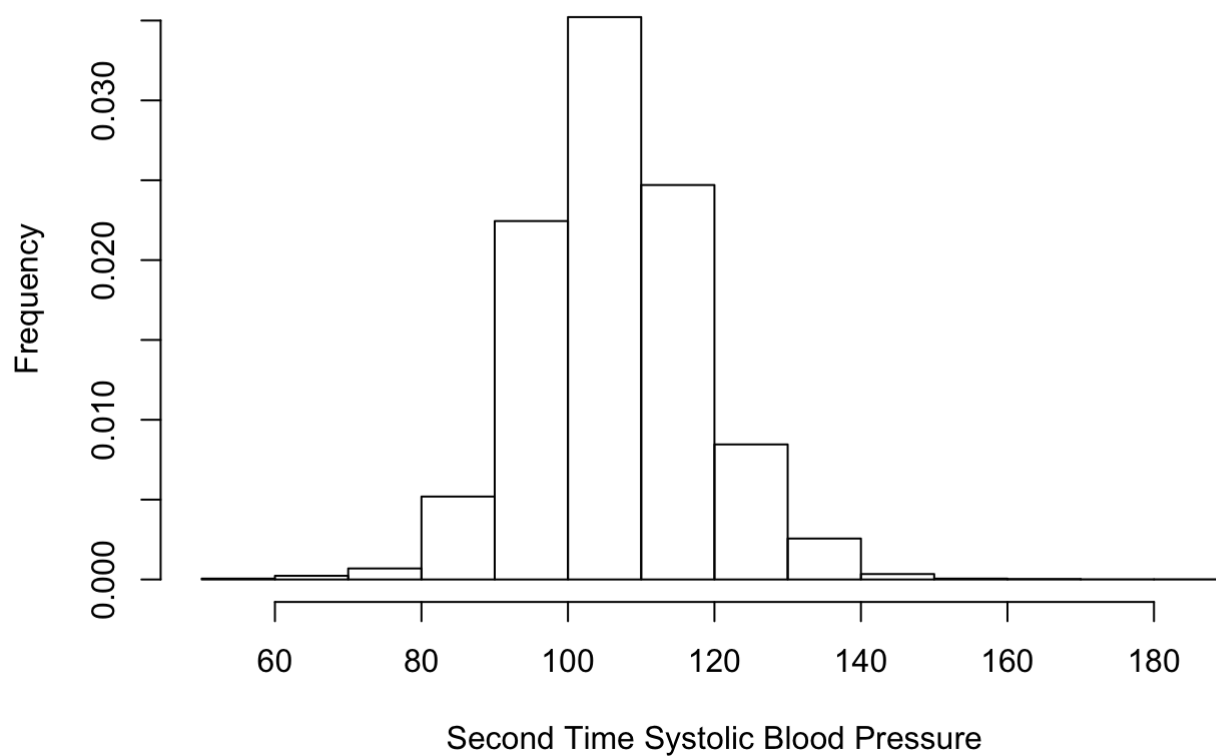


```
x2 <- hse.mk20$sys2om  
range(hse.mk20$sys2om, na.rm = T)
```

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

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

Distribution of Second Time Systolic Blood Pressure Measurement

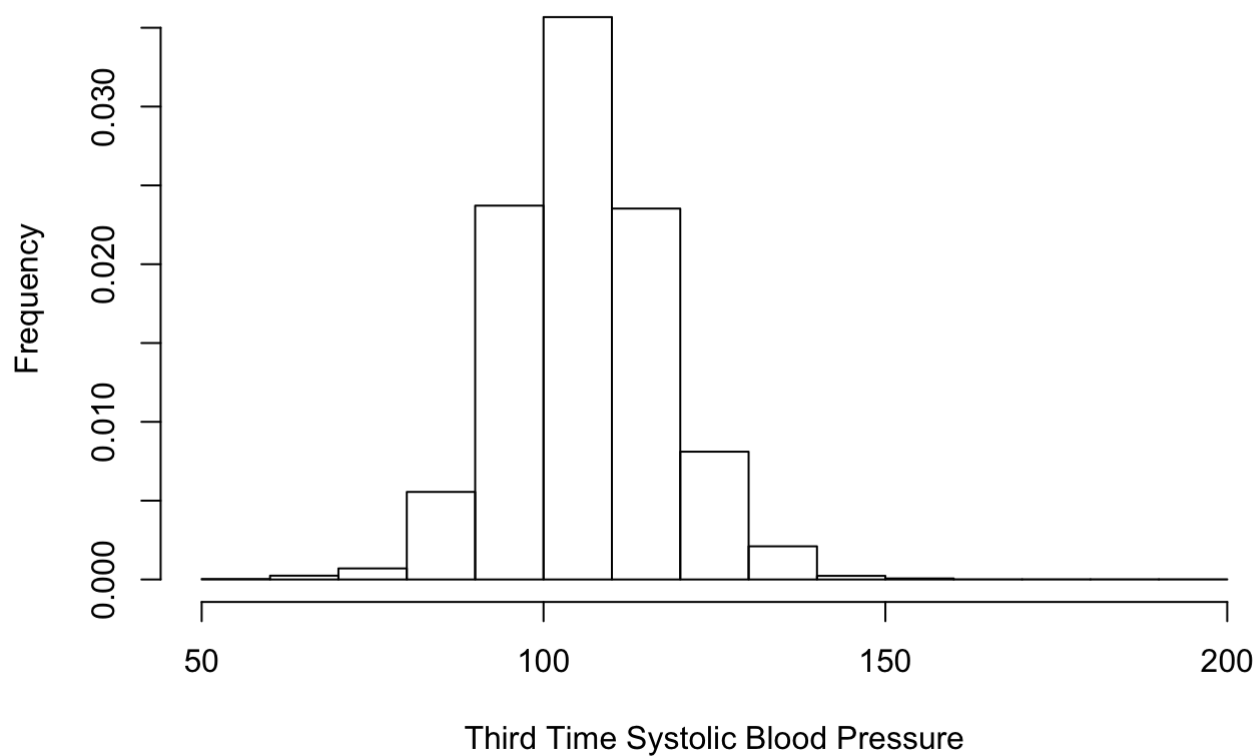


```
x3 <- hse.mk20$sys3om  
range(hse.mk20$sys3om, na.rm = T)
```

```
## [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

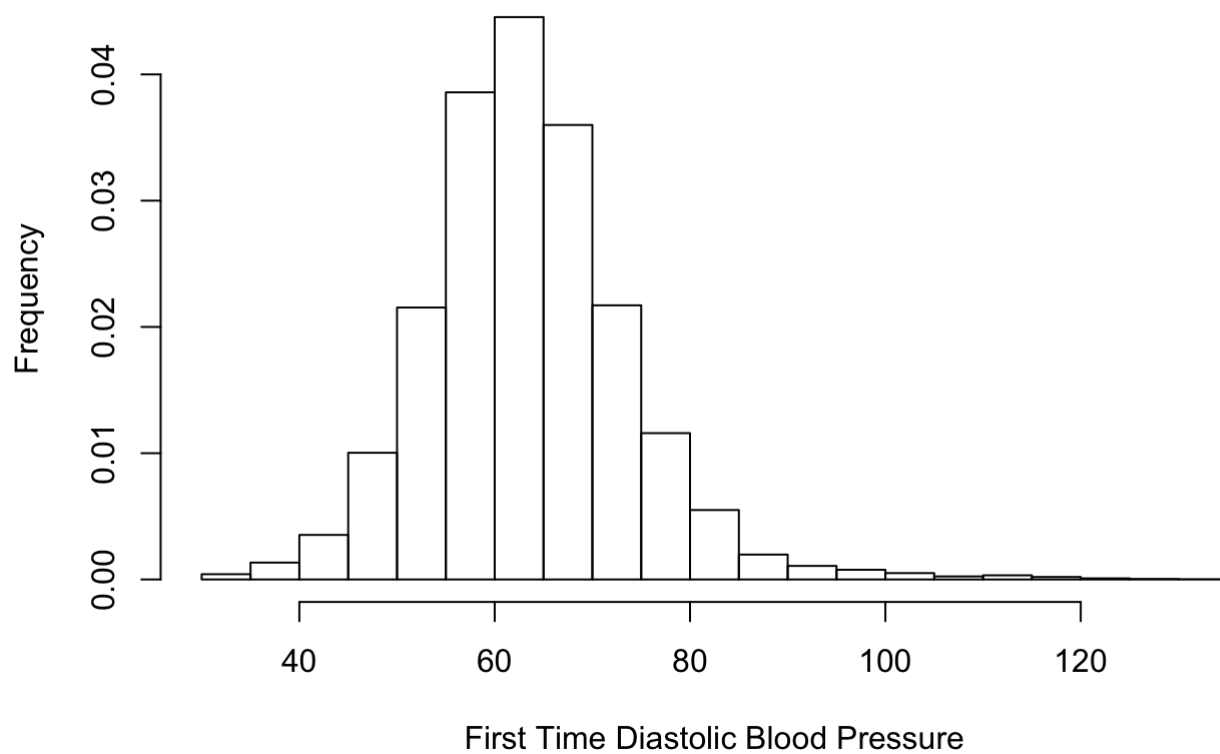


```
x4 <- hse.mk20$diaslom
range(hse.mk20$diaslom, na.rm = T)
```

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

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

Distribution of First Time Diastolic Blood Pressure Measurement

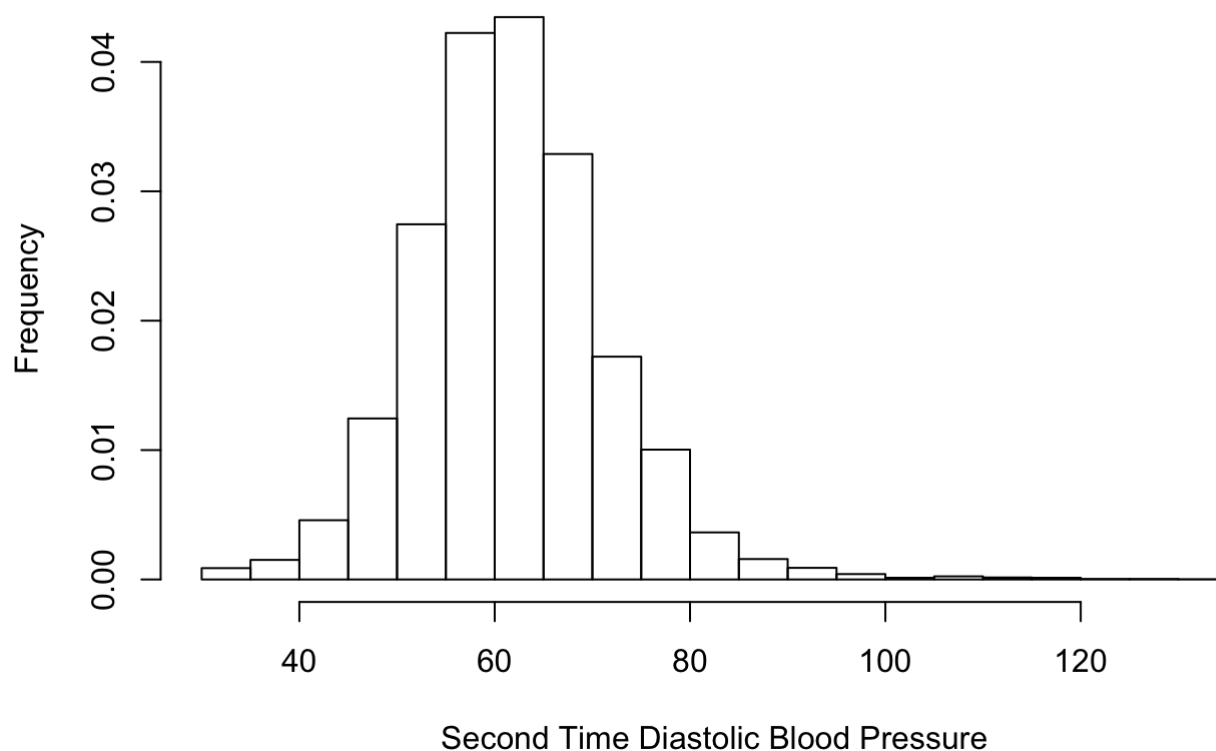


```
x5 <- hse.mk20$dias2om
range(hse.mk20$dias2om, na.rm = T)
```

```
## [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

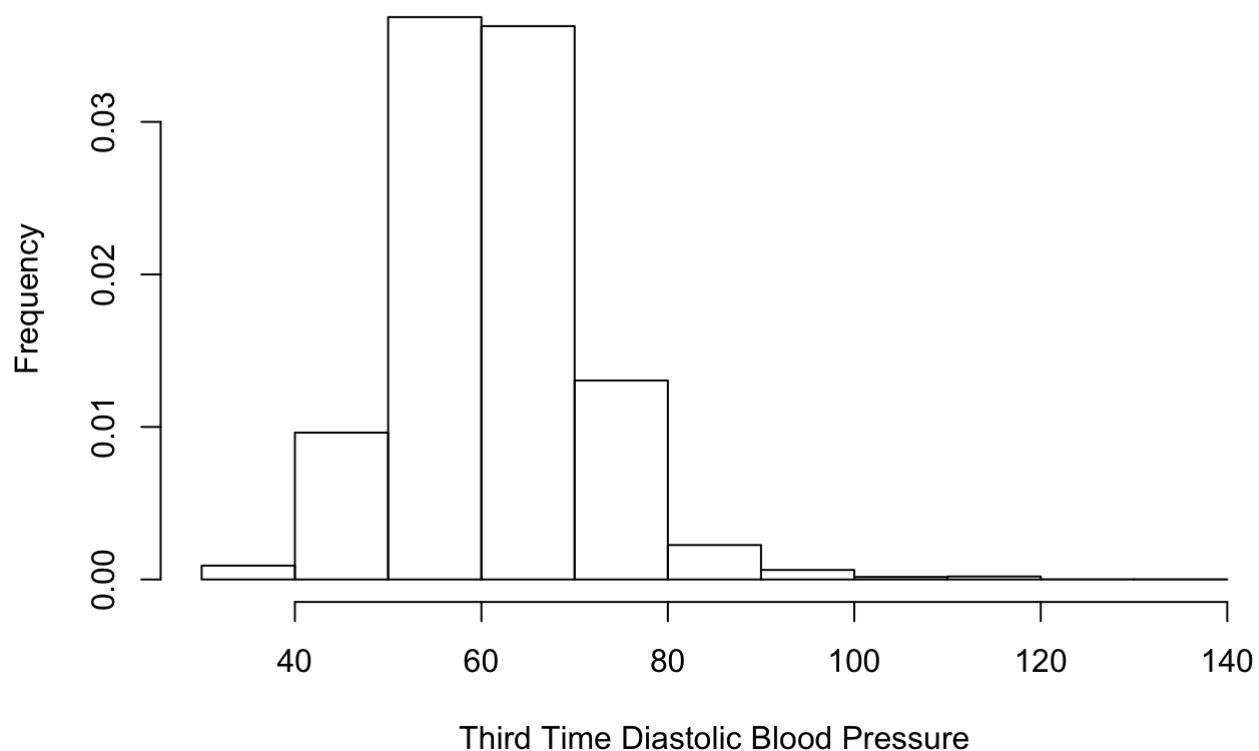


```
x6 <- hse.mk20$dias3om  
range(hse.mk20$dias3om, na.rm = T)
```

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

```
hist(x6, freq = FALSE, main = "Distribution of Third Time Diastolic Blood Pressure Me  
asurement"  
      , 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 #####
#####

hse.mk50 <- hse.mk20

# Multiple imputation
set.seed(1)
hse.mk50.mi <- aregImpute(~ hyper + sex + tenureb + origin + hhsize + addnum + imd +
  birthwt +
                        porftvg + gor + 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
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$tenureb <- hse.mk50.mi.r$tenureb
hse.mk60$origin <- hse.mk50.mi.r$origin
hse.mk60$birthwt <- hse.mk50.mi.r$birthwt
hse.mk60$porftvg <- hse.mk50.mi.r$porftvg
```

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
## 3834 3500
```

```
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(9)
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... 0.022286
## SE..... 0.0068587
## T-stat..... 3.2493
## p.val..... 0.001157
##
## Original number of observations..... 7334
## Original number of treated obs..... 3500
## Matched number of observations..... 3500
## Matched number of observations (unweighted). 3500
```

```
# 1 -- Estimate effect of deprivation on hypertention (significant)
# 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.....         1.5037              1.5037
## mean control.....           1.4877              1.492
## std mean diff.....          3.1942              2.3426
##
## mean raw eQQ diff.....       0.016              0.011714
## med  raw eQQ diff.....        0                  0
## max  raw eQQ diff.....        1                  1
##
## mean eCDF diff.....          0.0079865          0.0058571
## med  eCDF diff.....          0.0079865          0.0058571
## max  eCDF diff.....          0.015973          0.011714
##
## var ratio (Tr/Co).....       1.0006              1.0002
## T-test p-value.....          0.17182            0.32829
##
##
## ***** (V2) tenureb *****
##                               Before Matching      After Matching
## mean treatment.....           3                  3
## mean control.....            2.2074             3.0534
## std mean diff.....           72.057             -4.857
##
## mean raw eQQ diff.....       0.79486            0.091714
## med  raw eQQ diff.....        0                  0
## max  raw eQQ diff.....        2                  2
##
## mean eCDF diff.....          0.13247            0.015286
## med  eCDF diff.....          0.025672           0.0094286
## max  eCDF diff.....          0.37393            0.036286
##
## var ratio (Tr/Co).....       1.6377              0.99882
## T-test p-value.....          < 2.22e-16          0.00075071
## KS Bootstrap p-value..        < 2.22e-16          0.002
## KS Naive p-value.....        < 2.22e-16          0.019938
## KS Statistic.....            0.37393            0.036286
##
##
## ***** (V3) origin *****
##                               Before Matching      After Matching
## mean treatment.....           2.5954             2.5954
## mean control.....            1.5579             2.5931
## std mean diff.....           29.841             0.065741
##
## mean raw eQQ diff.....       1.0397              0.208
## med  raw eQQ diff.....        0                  0
## max  raw eQQ diff.....        8                  3
##
## mean eCDF diff.....          0.05764            0.011556
## med  eCDF diff.....          0.054904           0.008
## max  eCDF diff.....          0.17234            0.056286
##
## var ratio (Tr/Co).....       2.9274              0.88627
## T-test p-value.....          < 2.22e-16          0.97455
## KS Bootstrap p-value..        < 2.22e-16          < 2.22e-16

```

```

## KS Naive p-value..... < 2.22e-16          3.0581e-05
## KS Statistic.....          0.17234          0.056286
##
##
## ***** (V4) hhsize *****
##                               Before Matching      After Matching
## mean treatment.....          4.1154          4.1154
## mean control.....          4.0733          4.0854
## std mean diff.....          3.3495          2.3847
##
## mean raw eQQ diff.....      0.31371          0.21686
## med  raw eQQ diff.....          0          0
## max  raw eQQ diff.....          2          2
##
## mean eCDF diff.....          0.031313          0.021686
## med  eCDF diff.....          0.019654          0.012286
## max  eCDF diff.....          0.08692          0.057143
##
## var ratio (Tr/Co).....          1.8639          1.4864
## T-test p-value.....          0.10453          0.27388
## KS Bootstrap p-value.. < 2.22e-16          < 2.22e-16
## KS Naive p-value.....      1.9688e-12          2.176e-05
## KS Statistic.....          0.08692          0.057143
##
##
## ***** (V5) addnum *****
##                               Before Matching      After Matching
## mean treatment.....          16.173          16.173
## mean control.....          16.138          15.981
## std mean diff.....          0.29106          1.6216
##
## mean raw eQQ diff.....      0.23829          0.41086
## med  raw eQQ diff.....          0          0
## max  raw eQQ diff.....          7          7
##
## mean eCDF diff.....          0.004676          0.0082653
## med  eCDF diff.....          0.0033611          0.0074286
## max  eCDF diff.....          0.014439          0.02
##
## var ratio (Tr/Co).....          0.97605          0.96926
## T-test p-value.....          0.90149          0.49575
## KS Bootstrap p-value..          0.654          0.324
## KS Naive p-value.....          0.8401          0.48581
## KS Statistic.....          0.014439          0.02
##
##
## ***** (V6) birthwt *****
##                               Before Matching      After Matching
## mean treatment.....          3.273          3.273
## mean control.....          3.4102          3.2945
## std mean diff.....          -21.734          -3.4168
##
## mean raw eQQ diff.....      0.13688          0.036406
## med  raw eQQ diff.....          0.12          0.03
## max  raw eQQ diff.....          0.61          1.01
##
## mean eCDF diff.....          0.041907          0.0094731
## med  eCDF diff.....          0.03573          0.0074286

```

```

## max eCDF diff..... 0.10735 0.028286
##
## var ratio (Tr/Co)..... 1.124 0.94777
## T-test p-value..... < 2.22e-16 0.14283
## KS Bootstrap p-value.. < 2.22e-16 0.098
## KS Naive p-value..... < 2.22e-16 0.12156
## KS Statistic..... 0.10735 0.028286
##
##
## ***** (V7) porftvg *****
## Before Matching After Matching
## mean treatment..... 3.3503 3.3503
## mean control..... 3.7071 3.3551
## std mean diff..... -20.595 -0.28035
##
## mean raw eQQ diff..... 0.356 0.027143
## med raw eQQ diff..... 0 0
## max raw eQQ diff..... 2 1
##
## mean eCDF diff..... 0.050973 0.0038776
## med eCDF diff..... 0.046684 0.0028571
## max eCDF diff..... 0.082217 0.010286
##
## var ratio (Tr/Co)..... 1.1073 0.99687
## T-test p-value..... < 2.22e-16 0.90495
## KS Bootstrap p-value.. < 2.22e-16 0.754
## KS Naive p-value..... 3.6161e-11 0.99256
## KS Statistic..... 0.082217 0.010286
##
##
## ***** (V8) gor *****
## Before Matching After Matching
## mean treatment..... 4.5314 4.5314
## mean control..... 5.7019 4.6463
## std mean diff..... -45.935 -4.5076
##
## mean raw eQQ diff..... 1.1694 0.20971
## med raw eQQ diff..... 1 0
## max raw eQQ diff..... 2 1
##
## mean eCDF diff..... 0.13005 0.023302
## med eCDF diff..... 0.15121 0.022286
## max eCDF diff..... 0.20502 0.054857
##
## var ratio (Tr/Co)..... 1.0171 0.9681
## T-test p-value..... < 2.22e-16 0.030398
## KS Bootstrap p-value.. < 2.22e-16 < 2.22e-16
## KS Naive p-value..... < 2.22e-16 5.3308e-05
## KS Statistic..... 0.20502 0.054857
##
##
## ***** (V9) aggr *****
## Before Matching After Matching
## mean treatment..... 2.4651 2.4651
## mean control..... 2.4982 2.4669
## std mean diff..... -6.6214 -0.34364
##
## mean raw eQQ diff..... 0.032857 0.0017143

```

```
## med   raw eQQ diff.....      0      0
## max   raw eQQ diff.....      1      1
##
## mean  eCDF diff.....      0.016516      0.00085714
## med   eCDF diff.....      0.016516      0.00085714
## max   eCDF diff.....      0.033031      0.0017143
##
## var ratio (Tr/Co).....      0.99518      0.99953
## T-test p-value.....      0.0046814      0.88537
##
##
## ***** (V10) year *****
##                               Before Matching      After Matching
## mean treatment.....      9.5771      9.5771
## mean control.....      9.4567      9.6457
## std mean diff.....      3.6404      -2.0727
##
## mean raw eQQ diff.....      0.122      0.29371
## med   raw eQQ diff.....      0      0
## max   raw eQQ diff.....      2      2
##
## mean eCDF diff.....      0.01185      0.025571
## med   eCDF diff.....      0.0098061      0.029143
## max   eCDF diff.....      0.026924      0.052571
##
## var ratio (Tr/Co).....      1.0419      1.1387
## T-test p-value.....      0.11591      0.36114
## KS Bootstrap p-value..      0.058      < 2.22e-16
## KS Naive p-value.....      0.14088      0.0001259
## KS Statistic.....      0.026924      0.052571
##
##
## 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 gor year   Number(s): 3 4 8 10
```

```
# Regression with a matched dataset
# Create a new dataset after matching
hse.mk.FX <- rbind(hse.mk90[psm$index.control,],hse.mk90[psm$index.treated,])
summary(lm(hyper ~ imd,data=hse.mk.FX))
```

```
##
## Call:
## lm(formula = hyper ~ imd, data = hse.mk.FX)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.10257 -0.10257 -0.08029 -0.08029  0.91971
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.080286   0.004869   16.490 < 2e-16 ***
## imd          0.022286   0.006886    3.237  0.00122 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.288 on 6998 degrees of freedom
## Multiple R-squared:  0.001495,    Adjusted R-squared:  0.001352
## F-statistic: 10.48 on 1 and 6998 DF,  p-value: 0.001215
```

```
# 2 -- Balance our datasets for more accurate measure & conduct regression analysis
#           Estimate effect of deprivation on hypertension (significant)

summary(lm(hyper ~ imd + sex + tenureb + origin + hhsize + addnum + birthwt + porftvg
+ gor + aggr + year,data=hse.mk.FX))
```

```
##
## Call:
## lm(formula = hyper ~ imd + sex + tenureb + origin + hhsize +
##      addnum + birthwt + porftvg + gor + aggr + year, data = hse.mk.FX)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.16450 -0.10761 -0.08654 -0.06556  0.97932
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.1395696  0.0349176   3.997 6.48e-05 ***
## imd          0.0221899  0.0068732   3.228 0.001250 **
## sex         -0.0308265  0.0069319  -4.447 8.84e-06 ***
## tenureb      0.0057732  0.0031735   1.819 0.068928 .
## origin      -0.0013636  0.0009946  -1.371 0.170411
## hhsize       0.0040579  0.0030191   1.344 0.178965
## addnum      -0.0001641  0.0003052  -0.538 0.590886
## birthwt     -0.0018254  0.0054132  -0.337 0.735967
## porftvg     -0.0003412  0.0020389  -0.167 0.867110
## gor         -0.0034047  0.0013706  -2.484 0.013010 *
## aggr         0.0073019  0.0069384   1.052 0.292657
## year        -0.0037710  0.0011371  -3.316 0.000917 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2873 on 6988 degrees of freedom
## Multiple R-squared:  0.008242,    Adjusted R-squared:  0.006681
## F-statistic: 5.279 on 11 and 6988 DF,  p-value: 2.3e-08
```

```
# 3 -- However, if we control other variables, deprivation is not very important predictors
# but still significant in explaining hypertension disparities between areas with different deprivation value.

# 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")
tbl <- CreateTableOne(vars = myVars, data = hse.mk90.bn, factorVars = catVars)
tbl
```



```
##
##
##      Overall
##      n      3834
##      sex = 2 (%) 1870 (48.8)
##      tenureb (%)
##          1      449 (11.7)
##          2     2768 (72.2)
##          3        15 ( 0.4)
##          4      577 (15.0)
##          5        25 ( 0.7)
##      origin (%)
##          1     3407 (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     691 (18.0)
##          4    2049 (53.4)
##          5     733 (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 (0.60)
##      porftvg (%)
##          0      156 ( 4.1)
##          1      124 ( 3.2)
##          2     670 (17.5)
##          3     850 (22.2)
##          4     728 (19.0)
##          5     549 (14.3)
##          6     757 (19.7)
##      gor (%)
##          1      184 ( 4.8)
##          2     404 (10.5)
##          3     351 ( 9.2)
##          4     420 (11.0)
##          5     293 ( 7.6)
##          6     547 (14.3)
##          7     264 ( 6.9)
##          8     798 (20.8)
```

```
##          9          573 (14.9)
##  aggr = 3 (%)    1910 (49.8)
##  year (%)
##          5          362 ( 9.4)
##          6          662 (17.3)
##          7          317 ( 8.3)
##          8          613 (16.0)
##         10          337 ( 8.8)
##         11          325 ( 8.5)
##         12          297 ( 7.7)
##         13          330 ( 8.6)
##         14          312 ( 8.1)
##         15          279 ( 7.3)
```

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

```

##
## Overall
## n 3500
## sex = 2 (%) 1763 (50.4)
## tenureb (%)
## 1 234 ( 6.7)
## 2 1394 (39.8)
## 3 30 ( 0.9)
## 4 1823 (52.1)
## 5 18 ( 0.5)
## 6 1 ( 0.0)
## origin (%)
## 1 2507 (71.6)
## 2 117 ( 3.3)
## 3 260 ( 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 (%)
## 2 296 ( 8.5)
## 3 765 (21.9)
## 4 1310 (37.4)
## 5 711 (20.3)
## 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 (0.63)
## porftvg (%)
## 0 275 ( 7.9)
## 1 144 ( 4.1)
## 2 736 (21.0)
## 3 772 (22.1)
## 4 604 (17.3)
## 5 436 (12.5)
## 6 533 (15.2)
## gor (%)
## 1 424 (12.1)
## 2 642 (18.3)
## 3 435 (12.4)
## 4 322 ( 9.2)
## 5 365 (10.4)
## 6 208 ( 5.9)

```

```
##          7          570 (16.3)
##          8          310 ( 8.9)
##          9          224 ( 6.4)
##   aggr = 3 (%)      1628 (46.5)
##   year (%)
##          5          305 ( 8.7)
##          6          613 (17.5)
##          7          263 ( 7.5)
##          8          596 (17.0)
##         10          293 ( 8.4)
##         11          233 ( 6.7)
##         12          262 ( 7.5)
##         13          335 ( 9.6)
##         14          284 ( 8.1)
##         15          316 ( 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, factorVars = catVars)
tab3
```

```

##
## Overall
## n 7000
## sex = 2 (%) 3485 (49.8)
## tenureb (%)
## 1 341 ( 4.9)
## 2 2958 (42.3)
## 3 40 ( 0.6)
## 4 3496 (49.9)
## 5 164 ( 2.3)
## 6 1 ( 0.0)
## origin (%)
## 1 5211 (74.4)
## 2 156 ( 2.2)
## 3 416 ( 5.9)
## 4 261 ( 3.7)
## 5 64 ( 0.9)
## 6 49 ( 0.7)
## 7 83 ( 1.2)
## 8 93 ( 1.3)
## 9 119 ( 1.7)
## 10 59 ( 0.8)
## 11 49 ( 0.7)
## 12 29 ( 0.4)
## 13 168 ( 2.4)
## 14 111 ( 1.6)
## 15 48 ( 0.7)
## 16 54 ( 0.8)
## 17 4 ( 0.1)
## 18 26 ( 0.4)
## hhsize (%)
## 2 465 ( 6.6)
## 3 1457 (20.8)
## 4 3001 (42.9)
## 5 1390 (19.9)
## 6 459 ( 6.6)
## 7 150 ( 2.1)
## 8 64 ( 0.9)
## 9 9 ( 0.1)
## 10 4 ( 0.1)
## 11 1 ( 0.0)
## birthwt (mean (SD)) 3.28 (0.64)
## porftvg (%)
## 0 533 ( 7.6)
## 1 341 ( 4.9)
## 2 1426 (20.4)
## 3 1534 (21.9)
## 4 1219 (17.4)
## 5 884 (12.6)
## 6 1063 (15.2)
## gor (%)
## 1 844 (12.1)
## 2 1227 (17.5)
## 3 812 (11.6)
## 4 756 (10.8)
## 5 630 ( 9.0)
## 6 689 ( 9.8)

```

```
##      7      782 (11.2)
##      8      734 (10.5)
##      9      526 ( 7.5)
## aggr = 3 (%) 3262 (46.6)
## year (%)
##      5      591 ( 8.4)
##      6     1129 (16.1)
##      7      458 ( 6.5)
##      8     1243 (17.8)
##     10      670 ( 9.6)
##     11      564 ( 8.1)
##     12      616 ( 8.8)
##     13      640 ( 9.1)
##     14      550 ( 7.9)
##     15      539 ( 7.7)
```

```
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, factorVars = catVars)
tab4
```

```
##
## Overall
## n 7000
## sex = 2 (%) 3485 (49.8)
## tenureb (%)
## 1 341 ( 4.9)
## 2 2958 (42.3)
## 3 40 ( 0.6)
## 4 3496 (49.9)
## 5 164 ( 2.3)
## 6 1 ( 0.0)
## origin (%)
## 1 5211 (74.4)
## 2 156 ( 2.2)
## 3 416 ( 5.9)
## 4 261 ( 3.7)
## 5 64 ( 0.9)
## 6 49 ( 0.7)
## 7 83 ( 1.2)
## 8 93 ( 1.3)
## 9 119 ( 1.7)
## 10 59 ( 0.8)
## 11 49 ( 0.7)
## 12 29 ( 0.4)
## 13 168 ( 2.4)
## 14 111 ( 1.6)
## 15 48 ( 0.7)
## 16 54 ( 0.8)
## 17 4 ( 0.1)
## 18 26 ( 0.4)
## hhsize (%)
## 2 465 ( 6.6)
## 3 1457 (20.8)
## 4 3001 (42.9)
## 5 1390 (19.9)
## 6 459 ( 6.6)
## 7 150 ( 2.1)
## 8 64 ( 0.9)
## 9 9 ( 0.1)
## 10 4 ( 0.1)
## 11 1 ( 0.0)
## birthwt (mean (SD)) 3.28 (0.64)
## porftvg (%)
## 0 533 ( 7.6)
## 1 341 ( 4.9)
## 2 1426 (20.4)
## 3 1534 (21.9)
## 4 1219 (17.4)
## 5 884 (12.6)
## 6 1063 (15.2)
## gor (%)
## 1 844 (12.1)
## 2 1227 (17.5)
## 3 812 (11.6)
## 4 756 (10.8)
## 5 630 ( 9.0)
## 6 689 ( 9.8)
```

```
##          7          782 (11.2)
##          8          734 (10.5)
##          9          526 ( 7.5)
##   aggr = 3 (%)    3262 (46.6)
##   year (%)
##          5          591 ( 8.4)
##          6        1129 (16.1)
##          7          458 ( 6.5)
##          8        1243 (17.8)
##         10          670 ( 9.6)
##         11          564 ( 8.1)
##         12          616 ( 8.8)
##         13          640 ( 9.1)
##         14          550 ( 7.9)
##         15          539 ( 7.7)
```

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... 0.012049
## SE..... 0.007091
## T-stat..... 1.6991
## p.val..... 0.089297
##
## Original number of observations..... 7334
## Original number of treated obs..... 3500
## Matched number of observations..... 3500
## Matched number of observations (unweighted). 20087
```

```
# 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..... 1.5037      1.5037
## mean control..... 1.4877      1.5007
## std mean diff..... 3.1942      0.61003
##
## mean raw eQQ diff..... 0.016      0.0025887
## med raw eQQ diff..... 0      0
## max raw eQQ diff..... 1      1
##
## mean eCDF diff..... 0.0079865      0.0012944
## med eCDF diff..... 0.0079865      0.0012944
## max eCDF diff..... 0.015973      0.0025887
##
## var ratio (Tr/Co)..... 1.0006      0.99995
## T-test p-value..... 0.17182      0.7996
##
##
## ***** (V2) tenureb *****
##
## Before Matching      After Matching
## mean treatment..... 3      3
## mean control..... 2.2074      3.0358
## std mean diff..... 72.057      -3.251
##
## mean raw eQQ diff..... 0.79486      0.09429
## med raw eQQ diff..... 0      0
## max raw eQQ diff..... 2      2
##
## mean eCDF diff..... 0.13247      0.015715
## med eCDF diff..... 0.025672      0.014387
## max eCDF diff..... 0.37393      0.046548
##
## var ratio (Tr/Co)..... 1.6377      0.97841
## T-test p-value..... < 2.22e-16      0.026233
## KS Bootstrap p-value.. < 2.22e-16      < 2.22e-16
## KS Naive p-value..... < 2.22e-16      < 2.22e-16
## KS Statistic..... 0.37393      0.046548
##
##
## ***** (V3) origin *****
##
## Before Matching      After Matching
## mean treatment..... 2.5954      2.5954
## mean control..... 1.5579      2.6006
## std mean diff..... 29.841      -0.14787
##
## mean raw eQQ diff..... 1.0397      0.2062
## med raw eQQ diff..... 0      0
## max raw eQQ diff..... 8      3
##
## mean eCDF diff..... 0.05764      0.011456
## med eCDF diff..... 0.054904      0.0074177
## max eCDF diff..... 0.17234      0.067954
##
## var ratio (Tr/Co)..... 2.9274      0.88175
## T-test p-value..... < 2.22e-16      0.94322
## KS Bootstrap p-value.. < 2.22e-16      < 2.22e-16

```

```

## KS Naive p-value..... < 2.22e-16          < 2.22e-16
## KS Statistic.....      0.17234              0.067954
##
##
## ***** (V4) hhsize *****
##                               Before Matching      After Matching
## mean treatment.....      4.1154              4.1154
## mean control.....      4.0733              4.099
## std mean diff.....      3.3495              1.3024
##
## mean raw eQQ diff.....    0.31371            0.18519
## med  raw eQQ diff.....      0                0
## max  raw eQQ diff.....      2                2
##
## mean eCDF diff.....      0.031313           0.018519
## med  eCDF diff.....      0.019654           0.012769
## max  eCDF diff.....      0.08692            0.066162
##
## var ratio (Tr/Co).....    1.8639              1.456
## T-test p-value.....      0.10453            0.55249
## KS Bootstrap p-value.. < 2.22e-16          < 2.22e-16
## KS Naive p-value.....    1.9688e-12         < 2.22e-16
## KS Statistic.....      0.08692            0.066162
##
##
## ***** (V5) addnum *****
##                               Before Matching      After Matching
## mean treatment.....      16.173             16.173
## mean control.....      16.138             15.845
## std mean diff.....      0.29106            2.7792
##
## mean raw eQQ diff.....    0.23829            0.35033
## med  raw eQQ diff.....      0                0
## max  raw eQQ diff.....      7                7
##
## mean eCDF diff.....      0.004676           0.0074
## med  eCDF diff.....      0.0033611         0.0060238
## max  eCDF diff.....      0.014439           0.02529
##
## var ratio (Tr/Co).....    0.97605            0.99963
## T-test p-value.....      0.90149            0.23925
## KS Bootstrap p-value..    0.664             < 2.22e-16
## KS Naive p-value.....      0.8401           5.2664e-06
## KS Statistic.....      0.014439            0.02529
##
##
## ***** (V6) birthwt *****
##                               Before Matching      After Matching
## mean treatment.....      3.273              3.273
## mean control.....      3.4102              3.285
## std mean diff.....     -21.734             -1.9001
##
## mean raw eQQ diff.....    0.13688            0.02292
## med  raw eQQ diff.....      0.12             0.02
## max  raw eQQ diff.....      0.61             0.76
##
## mean eCDF diff.....      0.041907           0.005864
## med  eCDF diff.....      0.03573            0.0036342

```

```

## max eCDF diff..... 0.10735 0.027431
##
## var ratio (Tr/Co)..... 1.124 0.94607
## T-test p-value..... < 2.22e-16 0.41783
## KS Bootstrap p-value.. < 2.22e-16 < 2.22e-16
## KS Naive p-value..... < 2.22e-16 5.4572e-07
## KS Statistic..... 0.10735 0.027431
##
##
## ***** (V7) porftvg *****
## Before Matching After Matching
## mean treatment..... 3.3503 3.3503
## mean control..... 3.7071 3.3414
## std mean diff..... -20.595 0.51541
##
## mean raw eQQ diff..... 0.356 0.023149
## med raw eQQ diff..... 0 0
## max raw eQQ diff..... 2 1
##
## mean eCDF diff..... 0.050973 0.003307
## med eCDF diff..... 0.046684 0.0034848
## max eCDF diff..... 0.082217 0.005974
##
## var ratio (Tr/Co)..... 1.1073 0.97935
## T-test p-value..... < 2.22e-16 0.82628
## KS Bootstrap p-value.. < 2.22e-16 0.47
## KS Naive p-value..... 3.6161e-11 0.866
## KS Statistic..... 0.082217 0.005974
##
##
## ***** (V8) gor *****
## Before Matching After Matching
## mean treatment..... 4.5314 4.5314
## mean control..... 5.7019 4.638
## std mean diff..... -45.935 -4.1806
##
## mean raw eQQ diff..... 1.1694 0.16717
## med raw eQQ diff..... 1 0
## max raw eQQ diff..... 2 1
##
## mean eCDF diff..... 0.13005 0.018575
## med eCDF diff..... 0.15121 0.01384
## max eCDF diff..... 0.20502 0.043809
##
## var ratio (Tr/Co)..... 1.0171 0.97652
## T-test p-value..... < 2.22e-16 0.044839
## KS Bootstrap p-value.. < 2.22e-16 < 2.22e-16
## KS Naive p-value..... < 2.22e-16 < 2.22e-16
## KS Statistic..... 0.20502 0.043809
##
##
## ***** (V9) aggr *****
## Before Matching After Matching
## mean treatment..... 2.4651 2.4651
## mean control..... 2.4982 2.4581
## std mean diff..... -6.6214 1.4074
##
## mean raw eQQ diff..... 0.032857 0.0056255

```

```

## med   raw eQQ diff.....      0      0
## max   raw eQQ diff.....      1      1
##
## mean  eCDF diff.....      0.016516      0.0028128
## med   eCDF diff.....      0.016516      0.0028128
## max   eCDF diff.....      0.033031      0.0056255
##
## var ratio (Tr/Co).....      0.99518      1.0022
## T-test p-value.....      0.0046814      0.55567
##
##
## ***** (V10) year *****
##                               Before Matching      After Matching
## mean treatment.....      9.5771      9.5771
## mean control.....      9.4567      9.6538
## std mean diff.....      3.6404      -2.3165
##
## mean raw eQQ diff.....      0.122      0.19864
## med   raw eQQ diff.....      0      0
## max   raw eQQ diff.....      2      2
##
## mean eCDF diff.....      0.01185      0.01724
## med   eCDF diff.....      0.0098061      0.020287
## max   eCDF diff.....      0.026924      0.029322
##
## var ratio (Tr/Co).....      1.0419      1.1376
## T-test p-value.....      0.11591      0.31231
## KS Bootstrap p-value..      0.07      < 2.22e-16
## KS Naive p-value.....      0.14088      6.3149e-08
## KS Statistic.....      0.026924      0.029322
##
##
## Before Matching Minimum p.value: < 2.22e-16
## Variable Name(s): tenureb origin hhsiz birthwt porftvg gor   Number(s): 2 3 4 6 7
8
##
## After Matching Minimum p.value: < 2.22e-16
## Variable Name(s): tenureb origin hhsiz addnum birthwt gor year   Number(s): 2 3 4
5 6 8 10

```