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)
```

```
################################      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 converte
d to three age groups

hse05.pc <- c("sys1om","dias1om","sex","tenureb","age","ethinda","hhsize","addnum","i
md2004",
              "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("sys1om","dias1om","sex","tenureb","age","ethinda","hhsize","addnum","i
md2004",
              "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("sys1om","dias1om","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("sys1om","dias1om","sex","tenureb","age","origin","hhsize","addnum","qi
md",
              "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("sys1om","dias1om","sex","tenureb","age","origin","hhsize","addnum","IM
D2007",
              "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("sys1om","dias1om","sex","tenureb","age","origin","hhsize","addnum","im
d2007",
              "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("sys1om","dias1om","Sex","tenureb","Age","Origin","HHSize","addnum","qi
md",
              "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] <- "hhsize"
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("sys1om","dias1om","Sex","tenureb","Age","Origin","HHSize","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] <- "hhsize"
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","HHSize","Addnum","qi
md",
              "BirthWt","porftvg","gor1","SYS2OM","SYS3OM","DIAS2OM","DIAS3OM")
hse13.mk2 <- hse13[,hse13.pc]
colnames(hse13.mk2)[1] <- "sys1om"
colnames(hse13.mk2)[2] <- "dias1om"
colnames(hse13.mk2)[3] <- "sex"
colnames(hse13.mk2)[6] <- "origin"
colnames(hse13.mk2)[7] <- "hhsize"
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("SYS1OM","DIAS1OM","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] <- "hhsize"
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("SYS1OM","DIAS1OM","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] <- "hhsize"
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 datas
ets      ##############################

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 n
umber in the datasets
#       This step ensures those cases are marked as NAs rather than continuous number

colnames(hse.mk20)
```

```
##  [1] "sys1om"  "dias1om" "sex"     "tenureb" "origin"  "hhsize"  "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$hhsize)
```

```
## [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$sys1om)
```

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

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

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

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

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

```
range(hse.mk20$dias1om)
```

```
## [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 age from 2-4 does not have valid BP measurements, so they ar
e dropped from the analysis at this step

hse.mk20.pc <- -c(which(hse.mk20$sys1om < 0 | hse.mk20$sys1om > 200 |
                        hse.mk20$sys2om < 0 | hse.mk20$sys2om > 200 |
                        hse.mk20$sys3om < 0 | hse.mk20$sys3om > 200 |
                        hse.mk20$dias1om < 0 | hse.mk20$dias1om > 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] "sys1om"  "dias1om" "sex"     "tenureb" "origin"  "hhsize"  "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$hhsize)
```

```
## [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$sys1om, 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
```

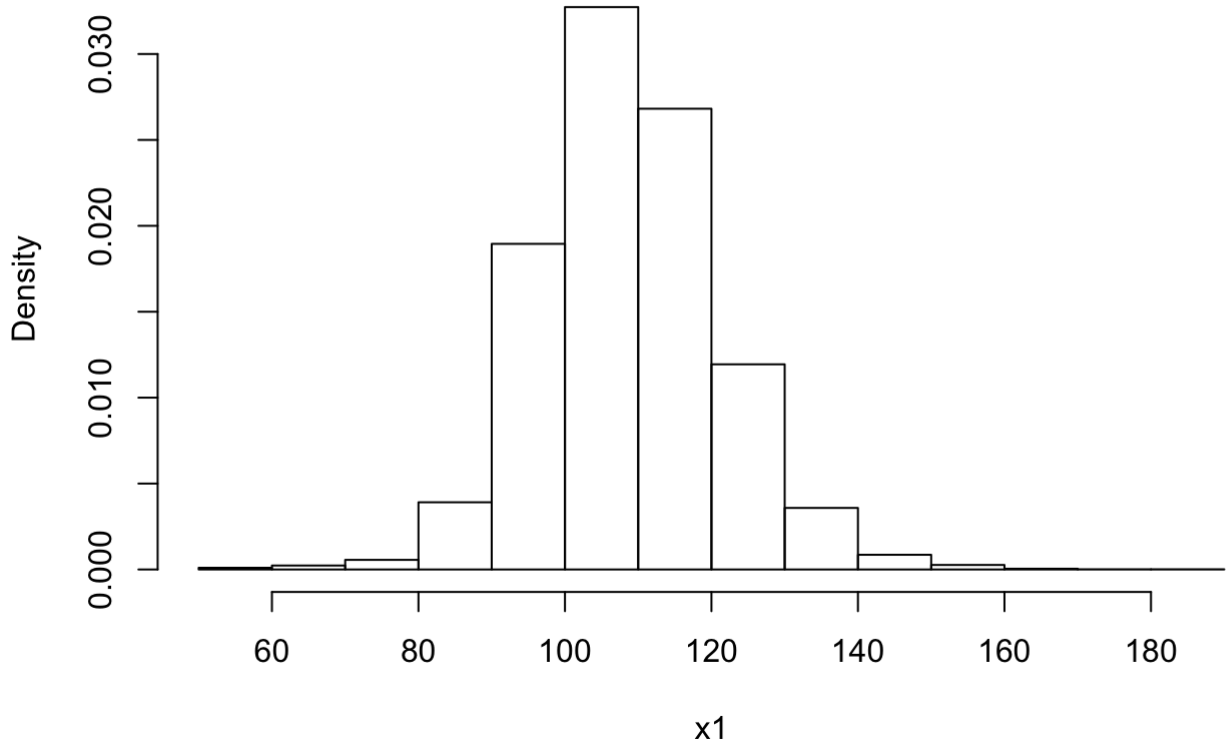
```
################################      Data Preparation  --  Average the BP for each ro
w      ################################

#       Looking into the distribution of BP measurements

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

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

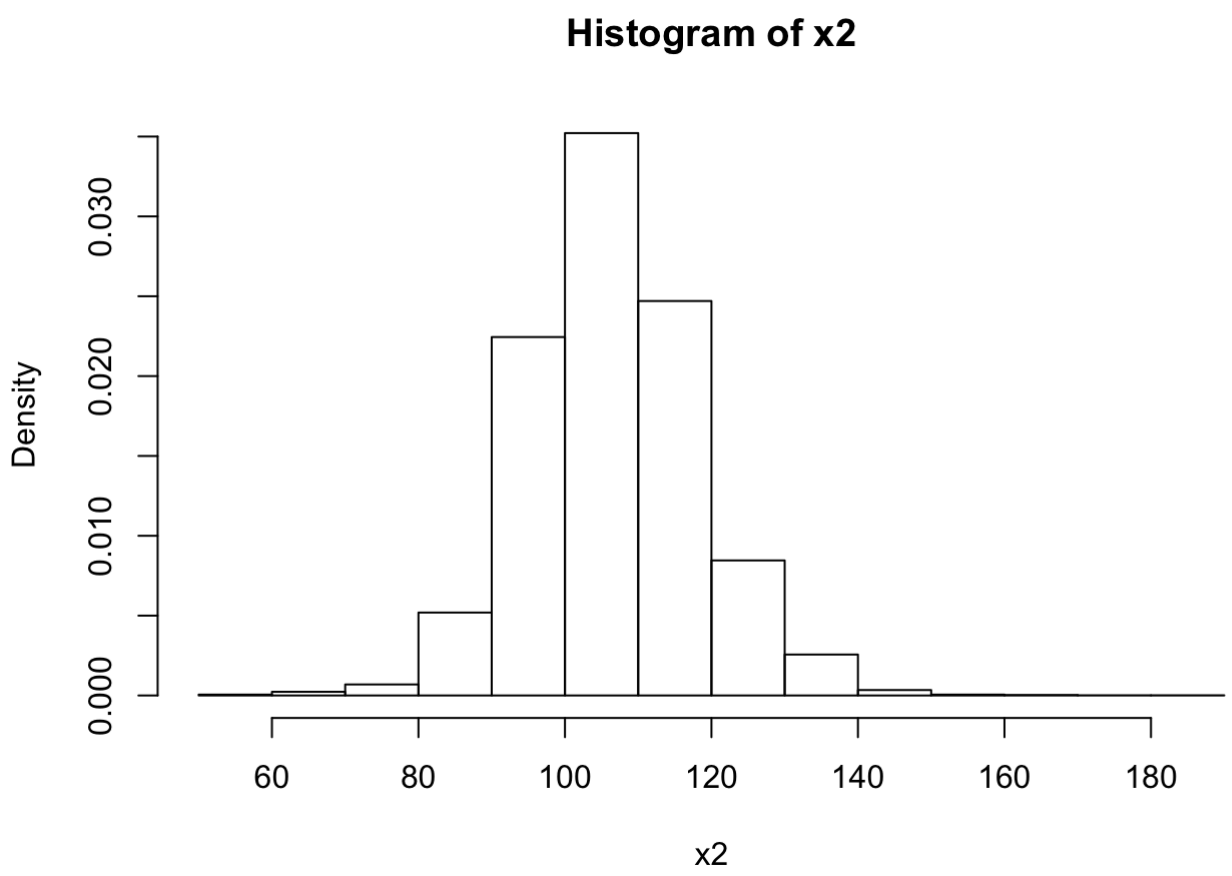
```
hist(x1, freq = FALSE)
```




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

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

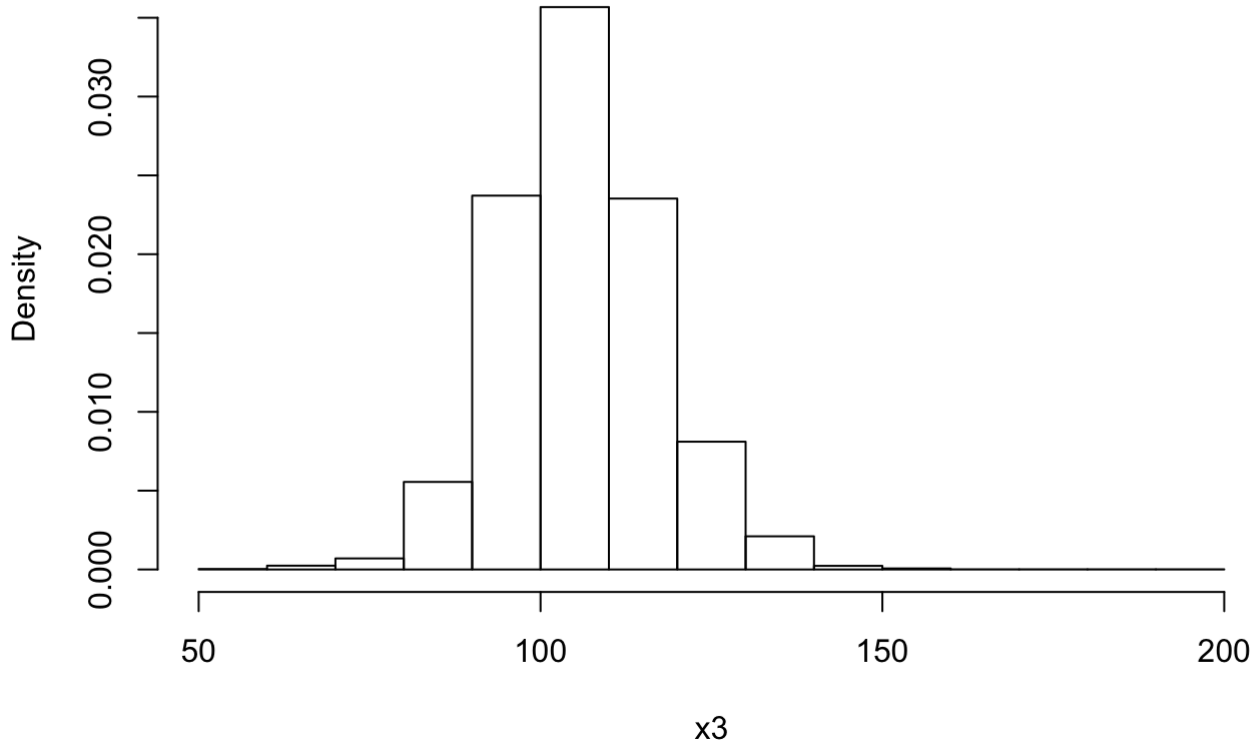
```
hist(x2, freq = FALSE)
```




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

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

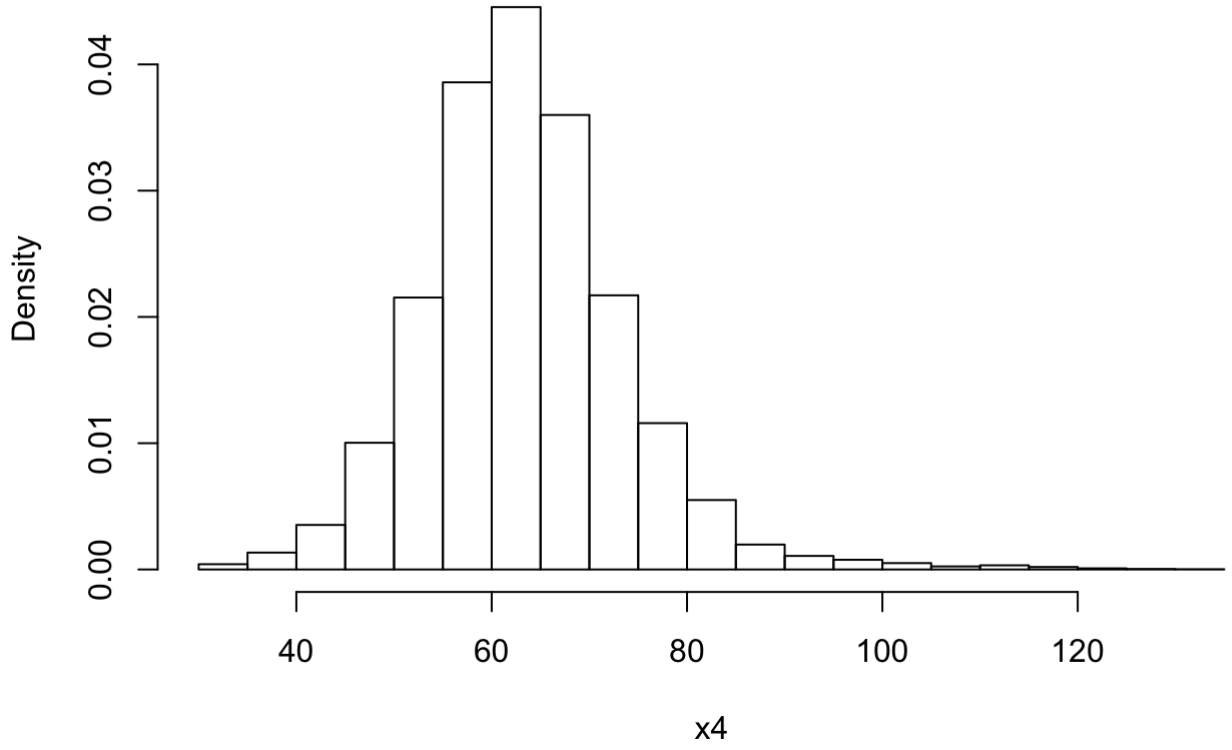
```
hist(x3, freq = FALSE)
```




```
x4 <- hse.mk20$dias1om
range(hse.mk20$dias1om, na.rm = T)
```

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

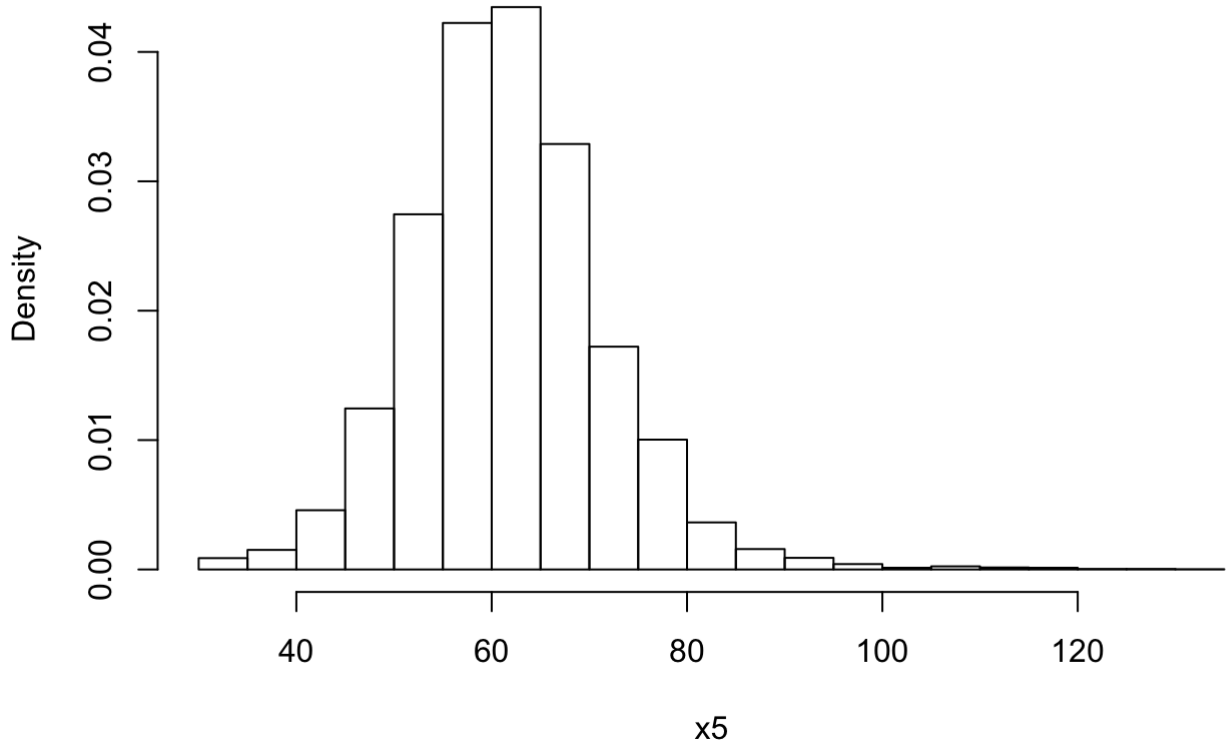
```
hist(x4, freq = FALSE)
```




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

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

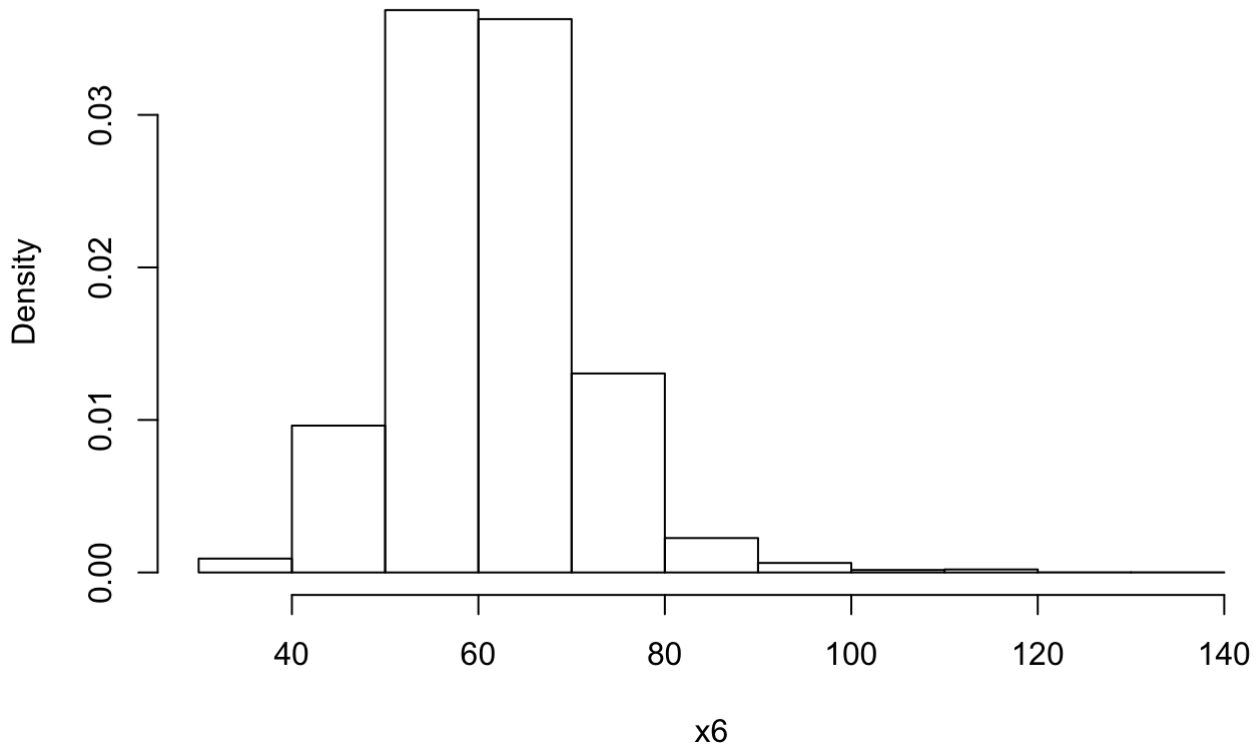
```
hist(x5, freq = FALSE)
```




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

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

```
hist(x6, freq = FALSE)
```




```
#       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$sys2om[i] + hse.mk20$sys3o
m[i]) / 3
      hse.mk20$diaavg[i] <- (hse.mk20$dias1om[i] + hse.mk20$dias2om[i] + hse.mk20$dia
s3om[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 = hs
e.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
```

```
####################   Data Preparation  --  Subset the hypertensive group for the cl
ustering analysis      ##################

#       hse.mk60 is the dataset for all children with regardless of whether they are
 hypertensive or not
#       hse.mk85 contains only the hypertensive children

hse.mk60.pc <- -c(which(hse.mk60$hyper == 0))
hse.mk85 <- hse.mk60[hse.mk60.pc,]

hse.mk85$hyper <- NULL
hse.mk85$sysavg <- NULL
hse.mk85$diaavg <- NULL
hse.mk85$sysavg <- NULL
hse.mk85$diaavg <- NULL
```

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, 1,
                       ifelse(hse.mk90$imd == 5, 0, NA))
table(hse.mk90$imd)
```

```
##
##    0    1
## 1830 2097
```

```
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(11)
matching.vars <- cbind(hse.mk90$fit.value)
psm <- Match(Y=hse.mk90$hyper, Tr=hse.mk90$imd, X=matching.vars, Weight = 2, ties = F
)
summary.Match(psm)
```

```
##
## Estimate...  0.018598
## SE.........  0.0075139
## T-stat.....  2.4751
## p.val......  0.013319
##
## Original number of observations..............  3927
## Original number of treated obs...............  2097
## Matched number of observations...............  2097
## Matched number of observations  (unweighted).  2097
```

```
#    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........     1.4831             1.4831
## mean control.........     1.4956             1.4649
## std mean diff.........    -2.5123             3.6254
##
## mean raw eQQ diff.....   0.012568           0.018121
## med  raw eQQ diff.....          0                  0
## max  raw eQQ diff.....          1                  1
##
## mean eCDF diff........  0.0062787          0.0090606
## med  eCDF diff........  0.0062787          0.0090606
## max  eCDF diff........   0.012557           0.018121
##
## var ratio (Tr/Co).....    0.99886             1.0038
## T-test p-value........    0.43243               0.24
##
##
## ***** (V2) tenureb *****
##                         Before Matching        After Matching
## mean treatment........     2.1497             2.1497
## mean control.........     3.1869             2.0482
## std mean diff.........    -128.24             12.559
##
## mean raw eQQ diff.....     1.0377            0.14354
## med  raw eQQ diff.....          1                  0
## max  raw eQQ diff.....          2                  2
##
## mean eCDF diff........    0.17286           0.028708
## med  eCDF diff........   0.031363          0.0090606
## max  eCDF diff........    0.49059            0.11683
##
## var ratio (Tr/Co).....     0.5675              0.816
## T-test p-value........ < 2.22e-16         5.8512e-08
## KS Bootstrap p-value.. < 2.22e-16         < 2.22e-16
## KS Naive p-value...... < 2.22e-16         7.4074e-13
## KS Statistic..........    0.49059            0.11683
##
##
## ***** (V3) origin *****
##                         Before Matching        After Matching
## mean treatment........     1.5131             1.5131
## mean control.........     2.8087             1.7787
## std mean diff.........    -67.155            -13.767
##
## mean raw eQQ diff.....     1.2984            0.28755
## med  raw eQQ diff.....          0                  0
## max  raw eQQ diff.....          9                  2
##
## mean eCDF diff........   0.076049           0.016859
## med  eCDF diff........   0.079276           0.003815
## max  eCDF diff........    0.21319            0.12732
##
## var ratio (Tr/Co).....    0.28059            0.87491
## T-test p-value........ < 2.22e-16         2.3668e-06
## KS Bootstrap p-value.. < 2.22e-16         < 2.22e-16
```

```
## KS Naive p-value..... < 2.22e-16         3.4417e-15
## KS Statistic.........    0.21319            0.12732
##
##
## ***** (V4) hhsize *****
##                         Before Matching        After Matching
## mean treatment........     4.0877             4.0877
## mean control.........     4.1743             4.1149
## std mean diff.........    -9.7108             -3.049
##
## mean raw eQQ diff.....    0.40546            0.19504
## med  raw eQQ diff.....          0                  0
## max  raw eQQ diff.....          2                  3
##
## mean eCDF diff........   0.040475           0.021671
## med  eCDF diff........   0.028262           0.015737
## max  eCDF diff........    0.10353           0.065808
##
## var ratio (Tr/Co).....    0.45237            0.55952
## T-test p-value........   0.018052            0.40808
## KS Bootstrap p-value.. < 2.22e-16         < 2.22e-16
## KS Naive p-value..... 1.5955e-09         0.00022749
## KS Statistic.........    0.10353           0.065808
##
##
## ***** (V5) addnum *****
##                         Before Matching        After Matching
## mean treatment........     16.147             16.147
## mean control.........     16.685             16.417
## std mean diff.........    -4.5391              -2.28
##
## mean raw eQQ diff.....    0.62459            0.79542
## med  raw eQQ diff.....          0                  1
## max  raw eQQ diff.....          5                  8
##
## mean eCDF diff........   0.014179           0.015794
## med  eCDF diff........   0.013992           0.015498
## max  eCDF diff........   0.037979           0.044826
##
## var ratio (Tr/Co).....    0.97447             1.0978
## T-test p-value........    0.15888            0.45047
## KS Bootstrap p-value..      0.064              0.028
## KS Naive p-value......    0.11931           0.029585
## KS Statistic.........   0.037979           0.044826
##
##
## ***** (V6) birthwt *****
##                         Before Matching        After Matching
## mean treatment........     3.4064             3.4064
## mean control.........     3.2368             3.3824
## std mean diff.........      28.26             4.0128
##
## mean raw eQQ diff.....    0.16895           0.076757
## med  raw eQQ diff.....       0.17               0.06
## max  raw eQQ diff.....       0.61                0.7
##
## mean eCDF diff........   0.055606           0.022989
## med  eCDF diff........   0.051557           0.018121
```

```
## max  eCDF diff........     0.1313           0.073915
##
## var ratio (Tr/Co).....    0.92313            0.98438
## T-test p-value........ < 2.22e-16            0.19158
## KS Bootstrap p-value.. < 2.22e-16         < 2.22e-16
## KS Naive p-value...... 4.6629e-15         2.1154e-05
## KS Statistic.........     0.1313           0.073915
##
##
## ***** (V7) porftvg *****
##                        Before Matching        After Matching
## mean treatment........     3.7492             3.7492
## mean control.........     3.2437             3.6371
## std mean diff.........     31.067              6.888
##
## mean raw eQQ diff.....    0.50437            0.11874
## med  raw eQQ diff.....          0                  0
## max  raw eQQ diff.....          2                  1
##
## mean eCDF diff........   0.072207           0.016963
## med  eCDF diff........   0.064757            0.01526
## max  eCDF diff........      0.116           0.047687
##
## var ratio (Tr/Co).....    0.87716            1.0186
## T-test p-value........ < 2.22e-16           0.020721
## KS Bootstrap p-value.. < 2.22e-16              0.002
## KS Naive p-value...... 7.5787e-12           0.016983
## KS Statistic.........      0.116           0.047687
##
##
## ***** (V8) gor *****
##                        Before Matching        After Matching
## mean treatment........      5.825              5.825
## mean control.........     4.0792             5.7129
## std mean diff.........     69.839             4.4831
##
## mean raw eQQ diff.....     1.7443            0.23891
## med  raw eQQ diff.....          2                  0
## max  raw eQQ diff.....          3                  1
##
## mean eCDF diff........    0.19397           0.026546
## med  eCDF diff........    0.22449           0.019075
## max  eCDF diff........    0.30776            0.07773
##
## var ratio (Tr/Co).....     1.0481              1.031
## T-test p-value........ < 2.22e-16           0.057349
## KS Bootstrap p-value.. < 2.22e-16         < 2.22e-16
## KS Naive p-value...... < 2.22e-16         6.2881e-06
## KS Statistic.........    0.30776            0.07773
##
##
## ***** (V9) aggr *****
##                        Before Matching        After Matching
## mean treatment........     2.4902             2.4902
## mean control.........     2.4672             2.4912
## std mean diff.........      4.602           -0.19074
##
## mean raw eQQ diff.....   0.022951          0.00095374
```

```
## med  raw eQQ diff.....          0                   0
## max  raw eQQ diff.....          1                   1
##
## mean eCDF diff........   0.011506         0.00047687
## med  eCDF diff........   0.011506         0.00047687
## max  eCDF diff........   0.023011         0.00095374
##
## var ratio (Tr/Co).....     1.0039            0.99993
## T-test p-value........    0.14993            0.95084
##
##
## ***** (V10) year *****
##                         Before Matching        After Matching
## mean treatment........     9.5031             9.5031
## mean control..........     9.7022              9.609
## std mean diff.........    -6.1378            -3.2638
##
## mean raw eQQ diff.....    0.20164            0.38817
## med  raw eQQ diff.....          0                   0
## max  raw eQQ diff.....          2                   2
##
## mean eCDF diff........   0.018896           0.034812
## med  eCDF diff........   0.018954           0.028135
## max  eCDF diff........   0.037739           0.086314
##
## var ratio (Tr/Co).....    0.95043            0.86366
## T-test p-value........   0.058498            0.31015
## KS Bootstrap p-value..      0.038         < 2.22e-16
## KS Naive p-value......     0.1236          3.282e-07
## KS Statistic..........   0.037739           0.086314
##
##
## 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): tenureb origin hhsize birthwt gor year  Number(s): 2 3 4 6 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
## -0.07058 -0.07058 -0.05198 -0.05198  0.94802
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.051979   0.005235   9.930   <2e-16 ***
## imd         0.018598   0.007403   2.512    0.012 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2397 on 4192 degrees of freedom
## Multiple R-squared:  0.001503,   Adjusted R-squared:  0.001265
## F-statistic: 6.311 on 1 and 4192 DF,  p-value: 0.01204
```

```
#    2  --   Balance our datasets for more accurate measure & conduct regression anal
ysis
#            Estimate effect of deprivation on hypertention (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
## -0.15217 -0.07680 -0.05710 -0.03777  0.99000
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.0741306  0.0386723   1.917  0.05532 .
## imd          0.0182801  0.0074174   2.464  0.01376 *
## sex         -0.0064419  0.0074358  -0.866  0.38635
## tenureb      0.0165997  0.0043746   3.795  0.00015 ***
## origin      -0.0001792  0.0018568  -0.096  0.92314
## hhsize      -0.0011908  0.0035517  -0.335  0.73744
## addnum       0.0005219  0.0003337   1.564  0.11787
## birthwt     -0.0032394  0.0061842  -0.524  0.60044
## porftvg     -0.0016013  0.0022934  -0.698  0.48509
## gor         -0.0075311  0.0015073  -4.997 6.07e-07 ***
## aggr         0.0006887  0.0074212   0.093  0.92607
## year         0.0008354  0.0011428   0.731  0.46483
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2388 on 4182 degrees of freedom
## Multiple R-squared:  0.01133,    Adjusted R-squared:  0.008727
## F-statistic: 4.356 on 11 and 4182 DF,  p-value: 1.619e-06
```

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

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 observ
ations.

# Propensity Scores Matching & Average Treatment Effect on Treated
set.seed(99)
matching.vars <- cbind(hse.mk90$fit.value)
psm <- Match(Y=hse.mk90$hyper, Tr=hse.mk90$imd, X=matching.vars, Weight = 2, ties = T
, version = "fast")
summary.Match(psm)
```

```
##
## Estimate...  0.022037
## SE.........  0.0073162
## T-stat.....  3.0121
## p.val......  0.0025945
##
## Original number of observations..............  3927
## Original number of treated obs...............  2097
## Matched number of observations...............  2097
## Matched number of observations  (unweighted).  5325
```

```
#    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........     1.4831             1.4831
## mean control.........     1.4956             1.4515
## std mean diff.........    -2.5123             6.3209
##
## mean raw eQQ diff.....   0.012568         0.00018779
## med  raw eQQ diff.....          0                  0
## max  raw eQQ diff.....          1                  1
##
## mean eCDF diff........  0.0062787         9.3897e-05
## med  eCDF diff........  0.0062787         9.3897e-05
## max  eCDF diff........   0.012557         0.00018779
##
## var ratio (Tr/Co).....    0.99886             1.0084
## T-test p-value........    0.43243           0.039439
##
##
## ***** (V2) tenureb *****
##                        Before Matching        After Matching
## mean treatment........     2.1497             2.1497
## mean control.........     3.1869             2.0636
## std mean diff.........    -128.24              10.65
##
## mean raw eQQ diff.....     1.0377            0.16977
## med  raw eQQ diff.....          1                  0
## max  raw eQQ diff.....          2                  2
##
## mean eCDF diff........    0.17286           0.028294
## med  eCDF diff........   0.031363            0.02216
## max  eCDF diff........    0.49059           0.093521
##
## var ratio (Tr/Co).....     0.5675            0.81271
## T-test p-value........ < 2.22e-16         4.1022e-06
## KS Bootstrap p-value.. < 2.22e-16         < 2.22e-16
## KS Naive p-value...... < 2.22e-16         < 2.22e-16
## KS Statistic..........    0.49059           0.093521
##
##
## ***** (V3) origin *****
##                        Before Matching        After Matching
## mean treatment........     1.5131             1.5131
## mean control.........     2.8087             1.7771
## std mean diff.........    -67.155            -13.682
##
## mean raw eQQ diff.....     1.2984             0.2539
## med  raw eQQ diff.....          0                  0
## max  raw eQQ diff.....          9                  3
##
## mean eCDF diff........   0.076049           0.014913
## med  eCDF diff........   0.079276          0.0060094
## max  eCDF diff........    0.21319            0.10423
##
## var ratio (Tr/Co).....    0.28059             0.8917
## T-test p-value........ < 2.22e-16         2.7715e-06
## KS Bootstrap p-value.. < 2.22e-16         < 2.22e-16
```

```
## KS Naive p-value..... < 2.22e-16         < 2.22e-16
## KS Statistic.........    0.21319            0.10423
##
##
## ***** (V4) hhsize *****
##                         Before Matching        After Matching
## mean treatment........     4.0877             4.0877
## mean control.........     4.1743             4.1248
## std mean diff.........    -9.7108            -4.1608
##
## mean raw eQQ diff.....    0.40546            0.23268
## med  raw eQQ diff.....          0                  0
## max  raw eQQ diff.....          2                  2
##
## mean eCDF diff........   0.040475           0.023268
## med  eCDF diff........   0.028262           0.020845
## max  eCDF diff........    0.10353           0.076056
##
## var ratio (Tr/Co).....    0.45237            0.52989
## T-test p-value........   0.018052            0.26507
## KS Bootstrap p-value.. < 2.22e-16         < 2.22e-16
## KS Naive p-value..... 1.5955e-09         8.3822e-14
## KS Statistic.........    0.10353           0.076056
##
##
## ***** (V5) addnum *****
##                         Before Matching        After Matching
## mean treatment........     16.147             16.147
## mean control.........     16.685             16.089
## std mean diff.........    -4.5391            0.49465
##
## mean raw eQQ diff.....    0.62459            0.85972
## med  raw eQQ diff.....          0                  1
## max  raw eQQ diff.....          5                  7
##
## mean eCDF diff........   0.014179           0.018046
## med  eCDF diff........   0.013992           0.012019
## max  eCDF diff........   0.037979           0.044883
##
## var ratio (Tr/Co).....    0.97447             1.1285
## T-test p-value........    0.15888            0.87062
## KS Bootstrap p-value..      0.078         < 2.22e-16
## KS Naive p-value......    0.11931         4.3891e-05
## KS Statistic.........   0.037979           0.044883
##
##
## ***** (V6) birthwt *****
##                         Before Matching        After Matching
## mean treatment........     3.4064             3.4064
## mean control.........     3.2368             3.3848
## std mean diff.........      28.26             3.6032
##
## mean raw eQQ diff.....    0.16895           0.055138
## med  raw eQQ diff.....       0.17               0.04
## max  raw eQQ diff.....       0.61               0.76
##
## mean eCDF diff........   0.055606           0.014829
## med  eCDF diff........   0.051557          0.0097653
```

```
## max  eCDF diff........     0.1313           0.048638
##
## var ratio (Tr/Co).....    0.92313            0.97909
## T-test p-value........ < 2.22e-16            0.24234
## KS Bootstrap p-value.. < 2.22e-16         < 2.22e-16
## KS Naive p-value...... 4.6629e-15         6.7618e-06
## KS Statistic.........     0.1313           0.048638
##
##
## ***** (V7) porftvg *****
##                        Before Matching        After Matching
## mean treatment........     3.7492             3.7492
## mean control.........     3.2437              3.656
## std mean diff.........     31.067             5.7275
##
## mean raw eQQ diff.....    0.50437           0.028169
## med  raw eQQ diff.....          0                  0
## max  raw eQQ diff.....          2                  1
##
## mean eCDF diff........   0.072207          0.0040241
## med  eCDF diff........   0.064757          0.0030047
## max  eCDF diff........      0.116          0.0088263
##
## var ratio (Tr/Co).....    0.87716            1.0176
## T-test p-value........ < 2.22e-16           0.050952
## KS Bootstrap p-value.. < 2.22e-16              0.704
## KS Naive p-value...... 7.5787e-12            0.98563
## KS Statistic.........      0.116          0.0088263
##
##
## ***** (V8) gor *****
##                        Before Matching        After Matching
## mean treatment........      5.825              5.825
## mean control.........     4.0792             5.7479
## std mean diff.........     69.839             3.0824
##
## mean raw eQQ diff.....     1.7443            0.22854
## med  raw eQQ diff.....          2                  0
## max  raw eQQ diff.....          3                  1
##
## mean eCDF diff........    0.19397           0.025394
## med  eCDF diff........    0.22449           0.021408
## max  eCDF diff........    0.30776            0.05784
##
## var ratio (Tr/Co).....     1.0481             1.0332
## T-test p-value........ < 2.22e-16            0.19287
## KS Bootstrap p-value.. < 2.22e-16         < 2.22e-16
## KS Naive p-value...... < 2.22e-16         3.6656e-08
## KS Statistic.........    0.30776            0.05784
##
##
## ***** (V9) aggr *****
##                        Before Matching        After Matching
## mean treatment........     2.4902             2.4902
## mean control.........     2.4672             2.4826
## std mean diff.........      4.602             1.5321
##
## mean raw eQQ diff.....   0.022951           0.021784
```

```
## med  raw eQQ diff.....          0                   0
## max  raw eQQ diff.....          1                   1
##
## mean eCDF diff........   0.011506           0.010892
## med  eCDF diff........   0.011506           0.010892
## max  eCDF diff........   0.023011           0.021784
##
## var ratio (Tr/Co).....     1.0039             1.0008
## T-test p-value........    0.14993            0.61901
##
##
## ***** (V10) year *****
##                         Before Matching        After Matching
## mean treatment........     9.5031             9.5031
## mean control..........     9.7022             9.6126
## std mean diff.........    -6.1378            -3.3751
##
## mean raw eQQ diff.....    0.20164             0.3493
## med  raw eQQ diff.....          0                  0
## max  raw eQQ diff.....          2                  2
##
## mean eCDF diff........   0.018896           0.029634
## med  eCDF diff........   0.018954           0.030235
## max  eCDF diff........   0.037739           0.058779
##
## var ratio (Tr/Co).....    0.95043            0.87529
## T-test p-value........   0.058498            0.29453
## KS Bootstrap p-value..      0.046         < 2.22e-16
## KS Naive p-value......     0.1236          2.046e-08
## KS Statistic..........   0.037739           0.058779
##
##
## 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): tenureb origin hhsize addnum birthwt gor year  Number(s): 2 3 4
5 6 8 10
```