

IHDP analyses for CH10 of GH

Data prep

```
cc <- read.table("data/ihdp.nlsy.imp1.txt", header = T, sep = "\t")
cc$treat <- cc$treat - 1

cc$ethnic = cc$hispanic
cc$ethnic[cc$black == 1] = 2
cc$ethnic[cc$white == 1] = 3

cc$educ = cc$lths
cc$educ[cc$hs == 1] = 2
cc$educ[cc$ltcoll == 1] = 3
cc$educ[cc$college == 1] = 4
cc$educ3 = cc$educ
cc$educ3[cc$educ > 2] = cc$educ3[cc$educ > 2] - 1

cc$bwg = (cc$bw > 2000) * 1

cc$state = cc$st5
cc$state[cc$st9 == 1] = 2
cc$state[cc$st12 == 1] = 3
cc$state[cc$st25 == 1] = 4
cc$state[cc$st36 == 1] = 5
cc$state[cc$st42 == 1] = 6
cc$state[cc$st48 == 1] = 7
cc$state[cc$st53 == 1] = 8

cc$state2 = cc$state
cc$state2[cc$st5 == 1] = 0

cc$state3 = cc$state
cc$state3[cc$st53 == 1] = 0

cc2 <- cc[cc$bw > 1500, ]

cc2$neg.bw = 2500 - cc2$bw
cc2$no.prenatal = 1 - cc2$prenatal
cc2$b.unmarr = 1 - cc2$b.marr

covs <- c("hispanic", "black", "white", "b.marr", "lths", "hs", "ltcoll", "college",
  "work.dur", "prenatal", "momage", "sex", "first", "preterm", "age", "dayskidh",
  "bw")
covs.st <- c("hispanic", "black", "white", "b.marr", "lths", "hs", "ltcoll", "college",
  "work.dur", "prenatal", "momage", "sex", "first", "age", "preterm", "dayskidh",
  "bw", "unemp.rt", "st5", "st9", "st12", "st25", "st36", "st42", "st48", "st53")

covs.nba <- c("hispanic", "black", "white", "b.marr", "lths", "hs", "ltcoll", "college",
  "work.dur", "prenatal", "momage", "sex", "first", "age")
covs.ba <- c("dayskidh", "bw", "preterm")
```

```

cc2$bwT = (cc2$bw - 1500)^2
# cc2$bwT = (cc2$neg.bw+5100)^2
cc2$dayskidT = log(cc2$dayskidh + 1)
cc2$pretermT = (cc2$preterm + 8)^2
cc2$momageT = (cc2$momage^2)

##### make a dataset to investigate bias amplification
data.ba = cc2[, c("ppvtr.36", "treat", covs.ba, covs.nba)]
lm(data.ba)

##
## Call:
## lm(formula = data.ba)
##
## Coefficients:
## (Intercept)      treat    dayskidh          bw      preterm
##  9.240e+01   1.169e+01  -1.396e-01   5.084e-04  -8.446e-03
##   hispanic      black      white      b.marr      lths
## -1.383e+01  -1.734e+01         NA   3.053e+00  -1.464e+01
##      hs      ltcoll    college    work.dur    prenatal
## -8.533e+00  -6.426e+00         NA   2.793e+00   4.299e+00
##    momage      sex      first      age
##  2.400e-02   1.181e+00   4.595e+00  -2.663e-02

data.nba = cc2[, c("ppvtr.36", "treat", covs.nba)]
lm(data.nba)

##
## Call:
## lm(formula = data.nba)
##
## Coefficients:
## (Intercept)      treat    hispanic      black      white
##  93.64649    9.56810  -13.81386  -17.38376         NA
##    b.marr      lths      hs      ltcoll    college
##  3.17895   -14.61671  -8.53332  -6.38220         NA
##   work.dur    prenatal    momage      sex      first
##  2.84278    4.48057    0.01170   1.16377   4.52370
##      age
##  -0.02874

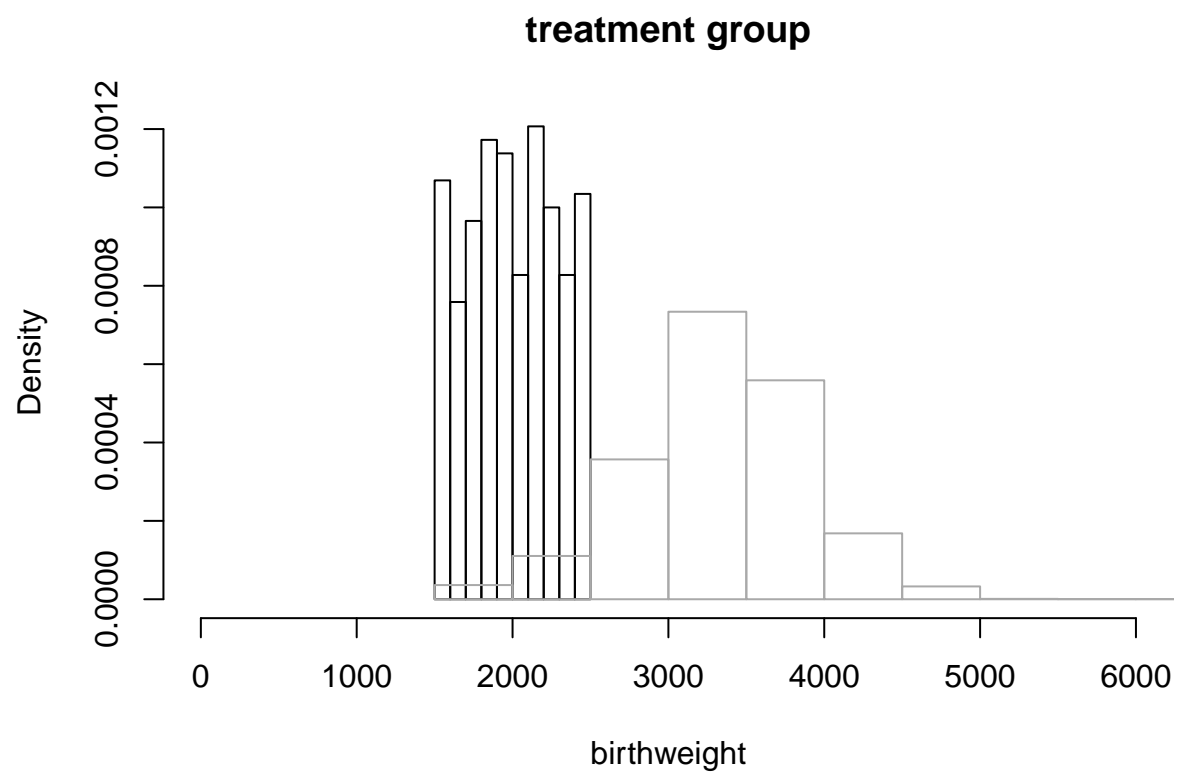
```

Univariate

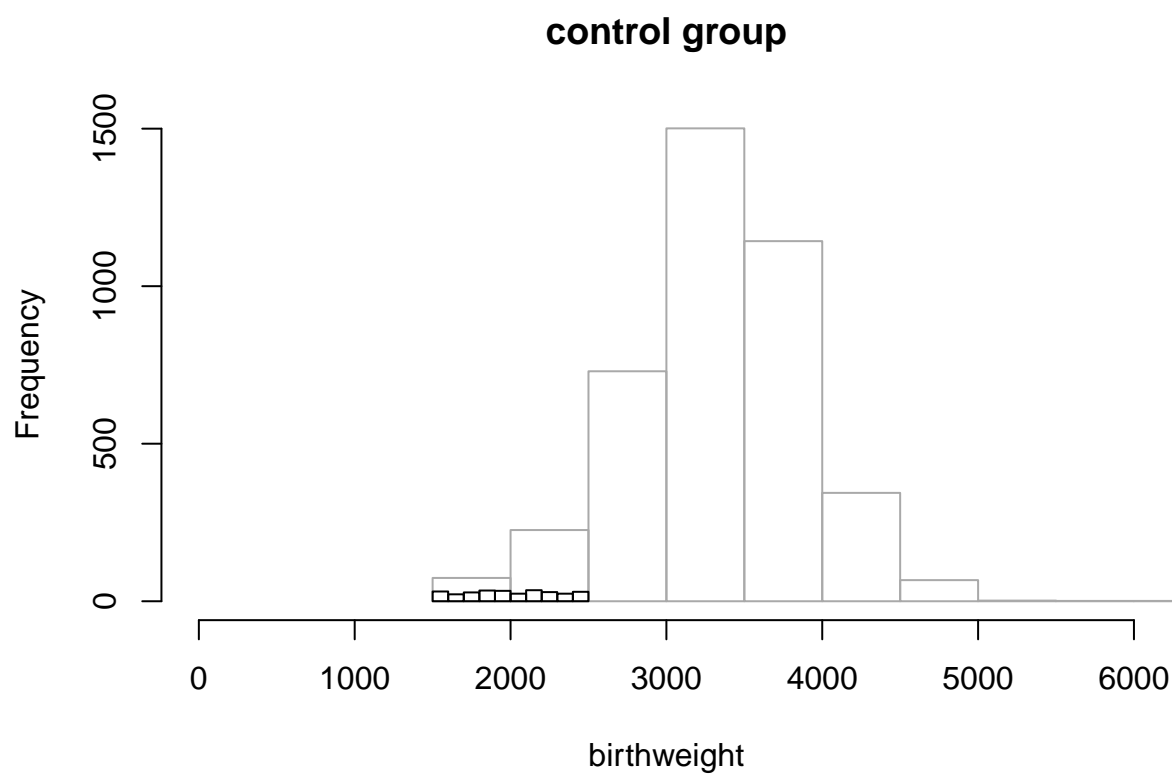
```

hist(cc2$bw[cc2$treat == 1], xlim = c(0, 6000), main = "treatment group", xlab = "birthweight",
     freq = F)
hist(cc2$bw[cc2$treat == 0], xlim = c(0, 6000), main = "control group", xlab = "birthweight",
     freq = F, add = T, border = "darkgrey")

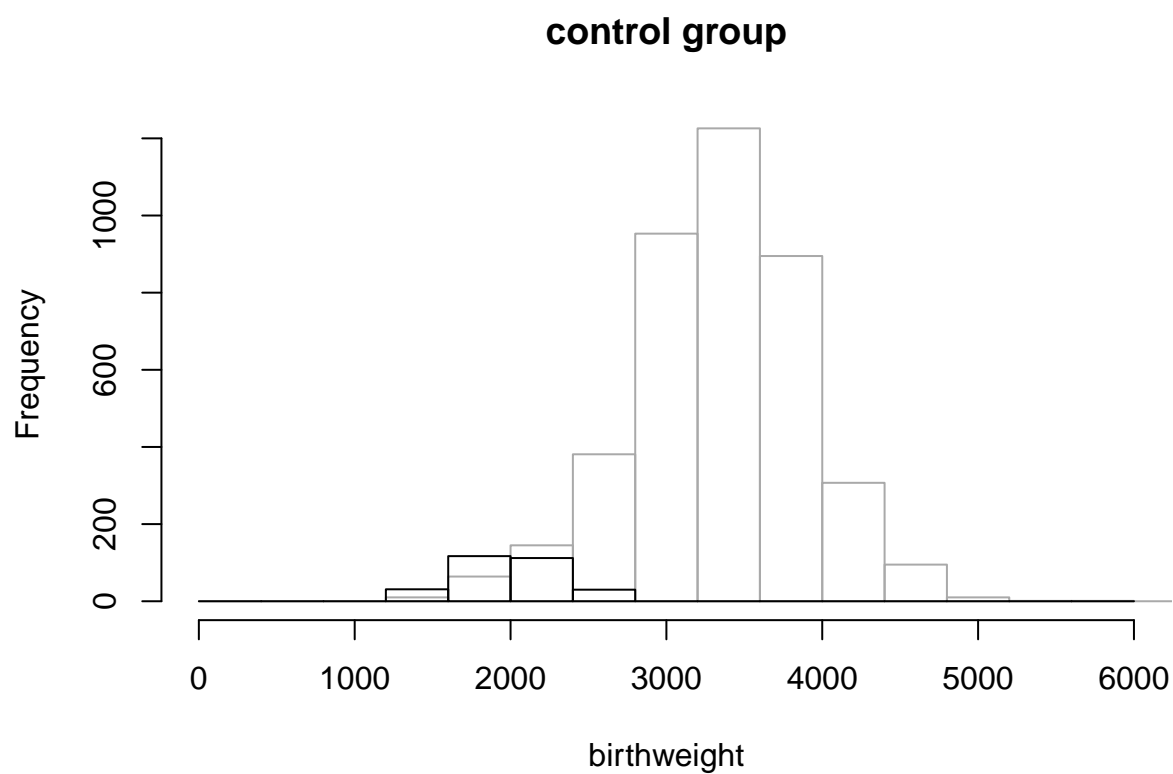
```



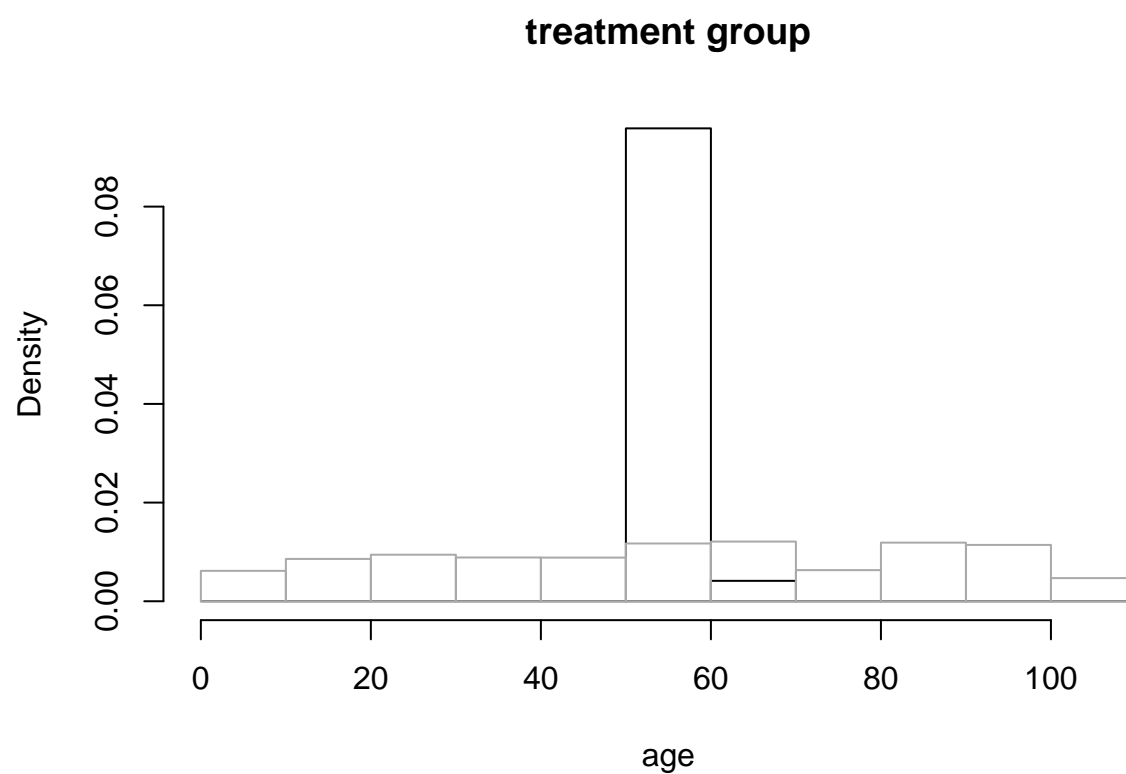
```
hist(cc2$bw[cc2$treat == 0], xlim = c(0, 6000), main = "control group", xlab = "birthweight",  
     border = "darkgrey")  
hist(cc2$bw[cc2$treat == 1], xlim = c(0, 6000), main = "treatment group", xlab = "birthweight",  
     add = T)
```



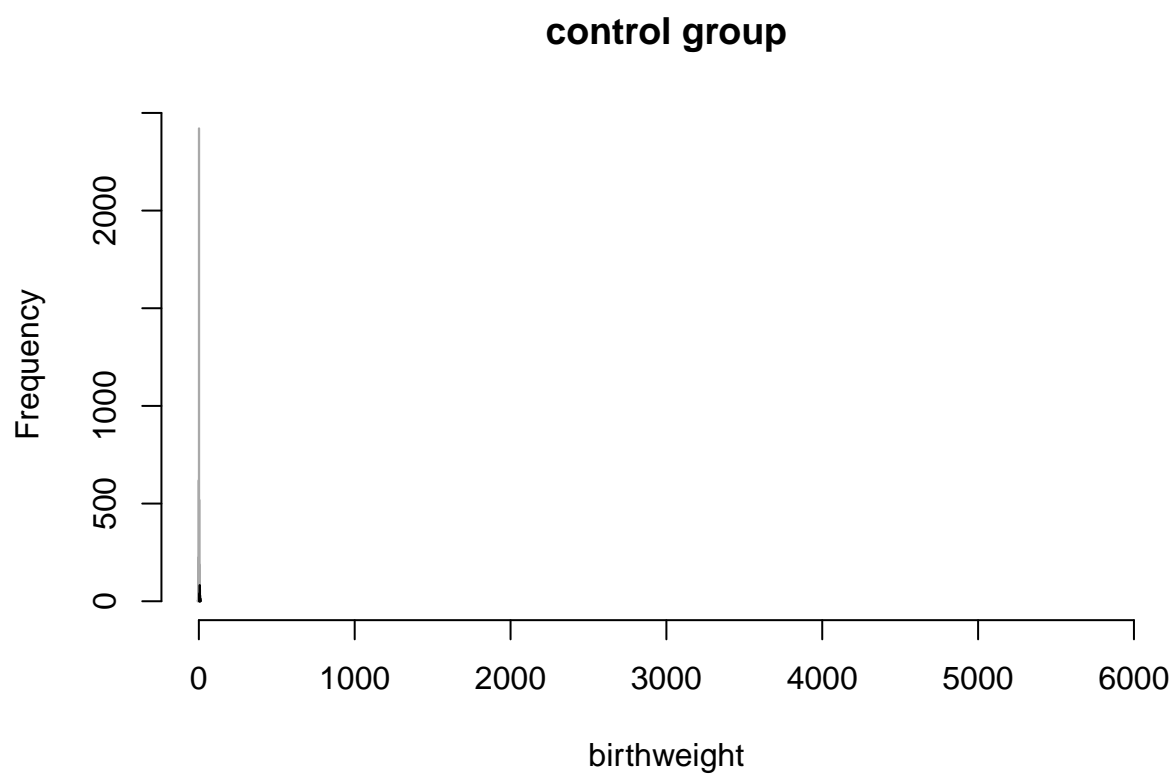
```
# breaks the same across groups
hist(cc2$bw[cc2$treat == 0], xlim = c(0, 6000), main = "control group", xlab = "birthweight",
     border = "darkgrey", breaks = seq(0, 8000, 400))
hist(cc2$bw[cc2$treat == 1], xlim = c(0, 6000), main = "treatment group", xlab = "birthweight",
     add = T, breaks = seq(0, 6000, 400))
```



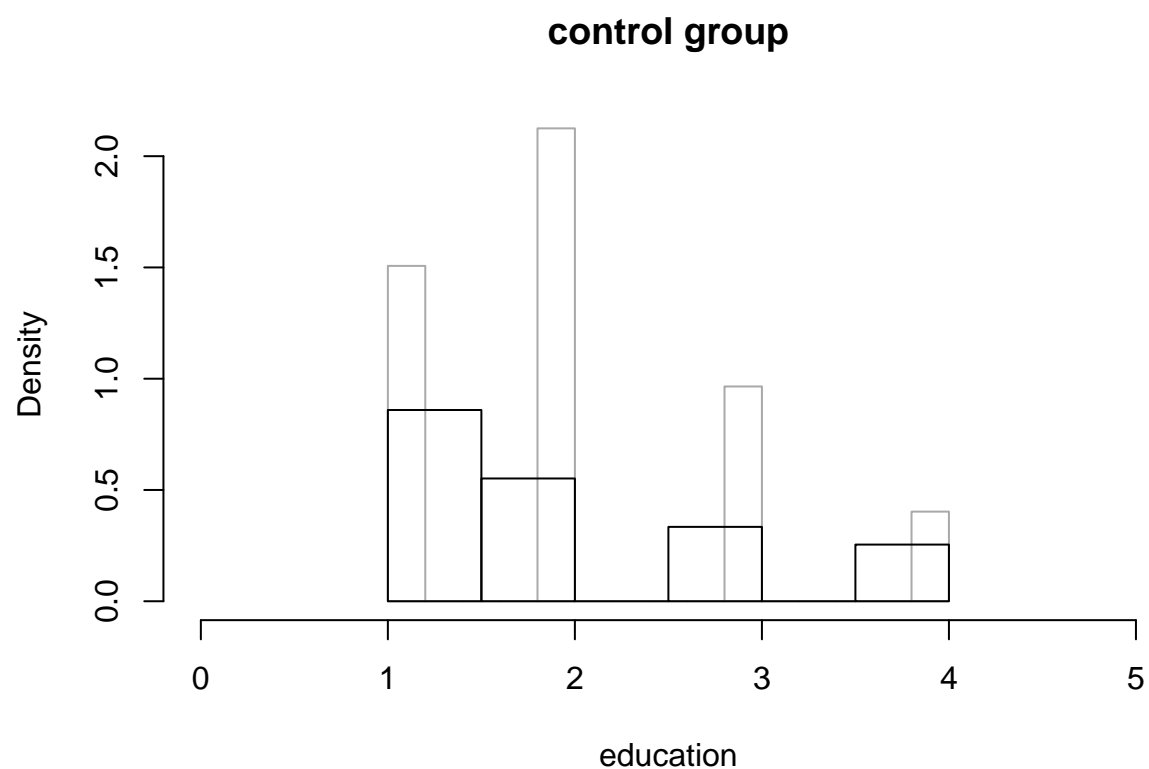
```
hist(cc2$age[cc2$treat == 1], xlim = c(0, 110), main = "treatment group", xlab = "age",  
     breaks = seq(0, 110, 10), freq = F)  
hist(cc2$age[cc2$treat == 0], xlim = c(0, 110), main = "control group", xlab = "age",  
     border = "darkgrey", add = T, breaks = seq(0, 110, 10), freq = F)
```



```
## now preterm
hist(cc2$preterm[cc2$treat == 0], xlim = c(0, 6000), main = "control group", xlab = "birthweight",
     border = "darkgrey")
hist(cc2$preterm[cc2$treat == 1], xlim = c(0, 6000), main = "treatment group", xlab = "birthweight",
     add = T)
```

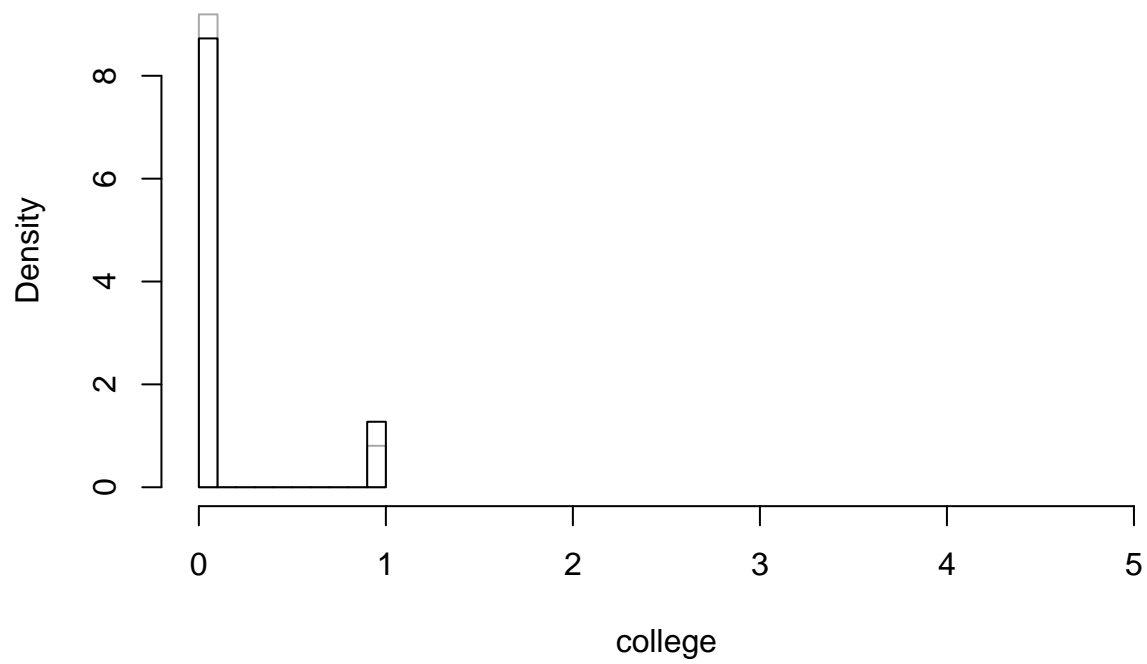


```
## now educ
hist(cc$educ[cc$treat == 0], xlim = c(0, 5), main = "control group", xlab = "education",
     border = "darkgrey", freq = F)
hist(cc$educ[cc$treat == 1], xlim = c(0, 5), main = "treatment group", xlab = "education",
     add = T, freq = F)
```



```
hist(cc$college[cc$treat == 0], xlim = c(0, 5), main = "control group", xlab = "college",  
     border = "darkgrey", freq = F)  
hist(cc$college[cc$treat == 1], xlim = c(0, 5), main = "treatment group", xlab = "college",  
     add = T, freq = F)
```


control group



```
hist(cc$educ3[cc$treat == 0], xlim = c(0, 5), main = "control group", xlab = "educ3",  
     border = "darkgrey", freq = F)  
hist(cc$educ3[cc$treat == 1], xlim = c(0, 5), main = "treatment group", xlab = "educ3",  
     add = T, freq = F)
```

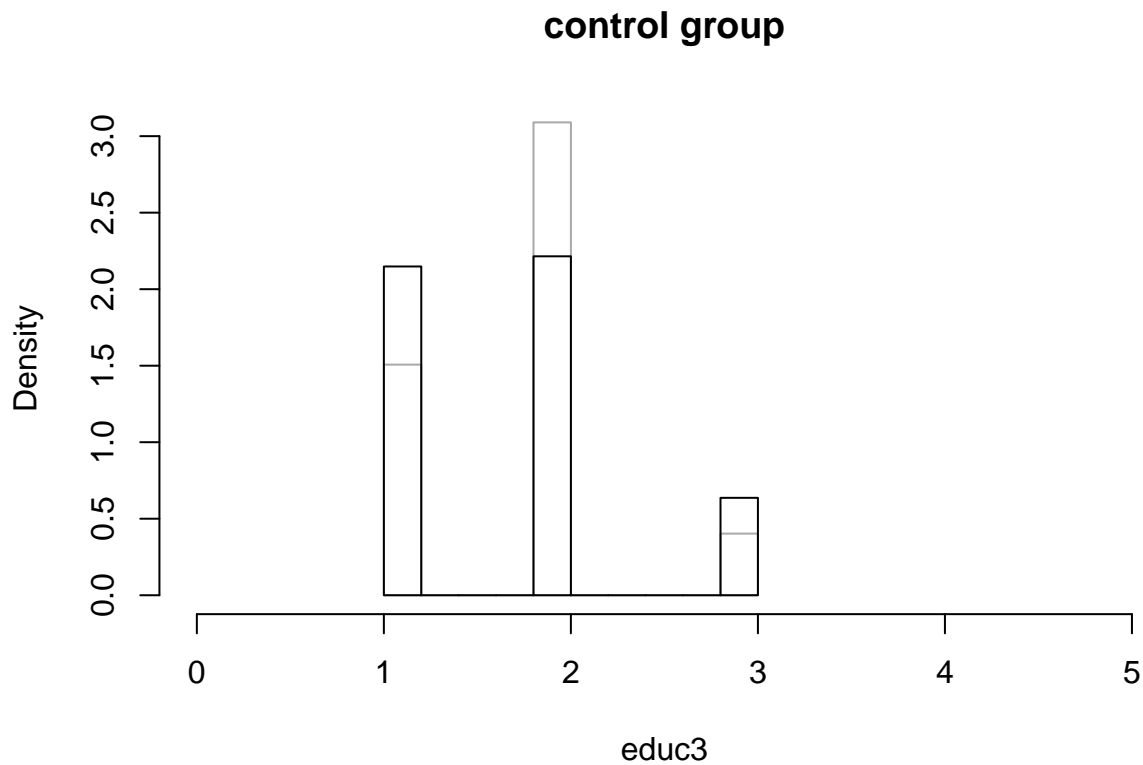
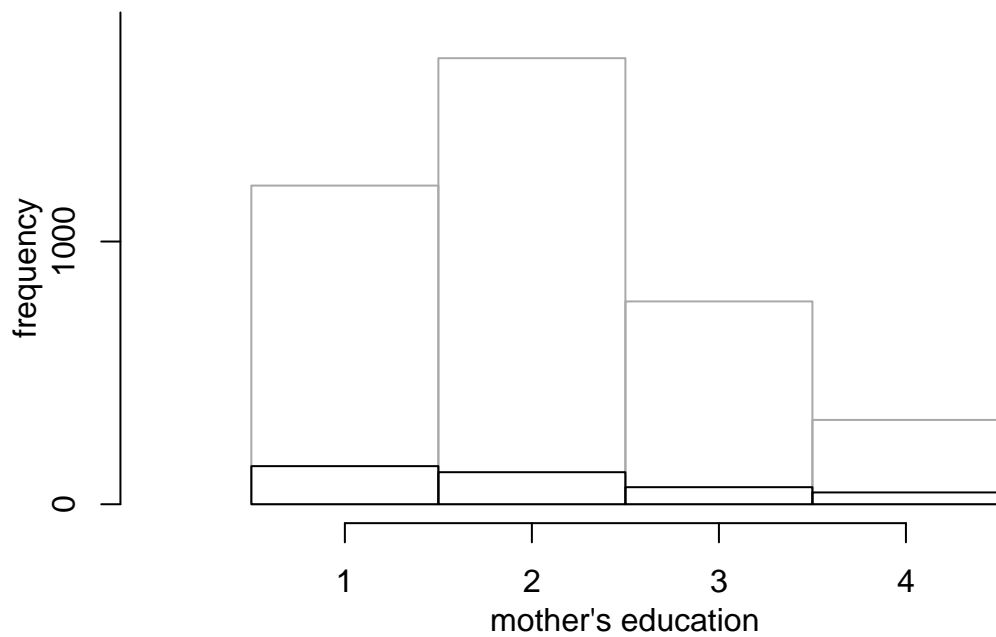


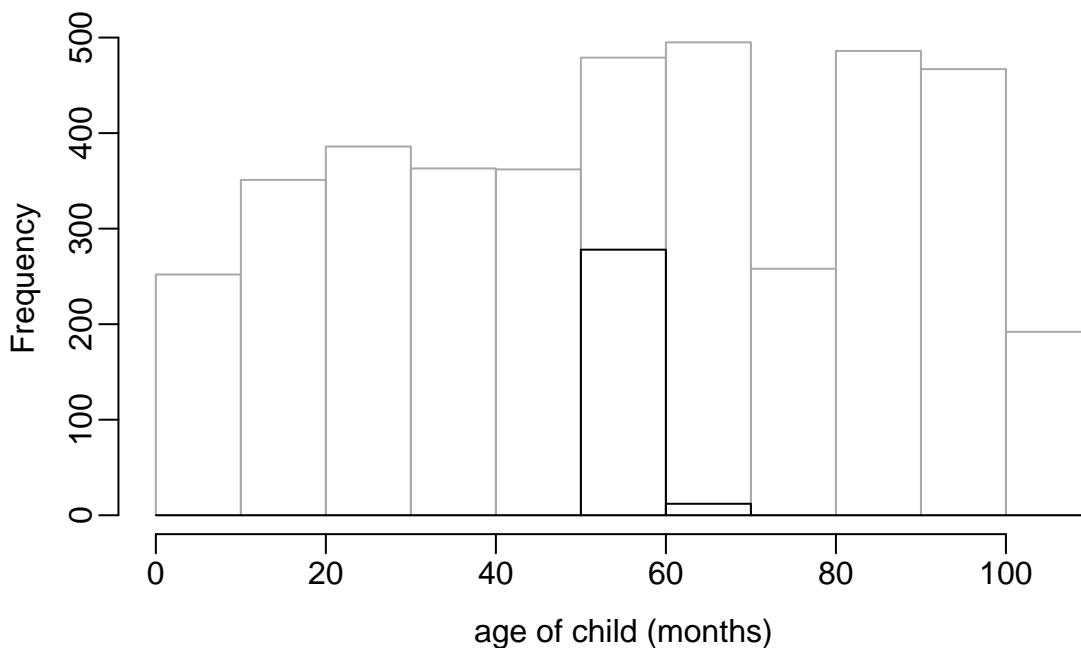
Figure 10.5

Matches text

```
## plots actually used to illustrate distinctions between overlap and balance now
## age
plot(x = seq(0, 5, 0.1), y = seq(0, 1800, (1800/50)), bty = "n", xaxt = "n", yaxt = "n",
     mgp = c(2, 0.5, 0), xlab = "mother's education", ylab = "frequency", type = "n",
     main = "")
axis(1, 1:4)
axis(2, c(0, 1000, 2000))
hist(cc2$educ[cc$treat == 0], xlim = c(0, 5), main = "", border = "darkgrey", breaks = c(0.5,
  1.5, 2.5, 3.5, 4.5), add = T)
hist(cc2$educ[cc$treat == 1], xlim = c(0, 5), xlab = "education", breaks = c(0.5,
  1.5, 2.5, 3.5, 4.5), add = T)
```



```
hist(cc2$age[cc2$treat == 0], xlim = c(0, 110), main = "", xlab = "age of child (months)",
     border = "darkgrey", breaks = seq(0, 110, 10), mgp = c(2, 0.5, 0))
hist(cc2$age[cc2$treat == 1], xlim = c(0, 110), xlab = "", breaks = seq(0, 110, 10),
     add = T)
```



```
# covs <-
# c('hispanic', 'black', 'white', 'b.marr', 'lths', 'hs', 'ltcoll', 'college', 'work.dur', 'prenatal', 'momage', 'neg.bw', 'preterm', 'dayskidh', 'sex', 'first', 'age', 'black', 'hispanic', 'white', 'b.unmarr', 'lths', 'hs', 'ltcoll', 'college', 'work.dur', 'no.prenatal', 'momage')
covs2 <- c("neg.bw", "preterm", "dayskidh", "sex", "first", "age", "black", "hispanic", "white", "b.unmarr", "lths", "hs", "ltcoll", "college", "work.dur", "no.prenatal", "momage")
cov.nms <- c("negative birth weight", "weeks preterm", "days in hospital", "male", "first born", "age", "black", "Hispanic", "white", "unmarried at birth", "less than high school", "high school graduate", "some college", "college graduate", "worked during pregnancy", "had no prenatal care", "age at birth")

diff.means = matrix(0, length(covs2), 6)
for (i in 1:length(covs2)) {
  diff.means[i, 1:2] <- c(mean(cc2[cc2$treat == 1, covs2[i]]), mean(cc2[cc2$treat == 0, covs2[i]]))
  diff.means[i, 3] <- diff.means[i, 1] - diff.means[i, 2]
  diff.means[i, 5] <- sqrt(var(cc2[cc2$treat == 1, covs2[i]])/sum(cc2$treat == 1) + var(cc2[cc2$treat == 0, covs2[i]])/sum(cc2$treat == 0))
  diff.means[i, 6] <- sqrt((var(cc2[cc2$treat == 1, covs2[i]]) + var(cc2[cc2$treat == 0, covs2[i]]))/2)
  diff.means[i, 4] <- diff.means[i, 3]/diff.means[i, 6]
}
dimnames(diff.means) <- list(covs2, c("treat", "control", "diff", "diff.std", "se", "sd"))
round(diff.means, 2)
```

```
##          treat control    diff diff.std    se    sd
```

## neg.bw	491.35	-835.27	1326.62	2.98	18.81	444.70
## preterm	6.07	1.18	4.89	2.49	0.12	1.97
## dayskidh	14.69	4.17	10.52	1.19	0.67	8.86
## sex	0.51	0.50	0.01	0.02	0.03	0.50
## first	0.48	0.42	0.06	0.12	0.03	0.50
## age	56.67	56.35	0.32	0.02	0.47	20.59
## black	0.50	0.28	0.22	0.46	0.03	0.48
## hispanic	0.09	0.21	-0.12	-0.34	0.02	0.36
## white	0.40	0.50	-0.10	-0.20	0.03	0.50
## b.unmarr	0.57	0.31	0.26	0.53	0.03	0.48
## lths	0.43	0.30	0.13	0.28	0.03	0.48
## hs	0.28	0.42	-0.14	-0.30	0.03	0.47
## ltcoll	0.17	0.19	-0.03	-0.07	0.02	0.38
## college	0.12	0.08	0.04	0.12	0.02	0.30
## work.dur	0.59	0.62	-0.03	-0.06	0.03	0.49
## no.prenatal	0.04	0.01	0.03	0.19	0.01	0.17
## momage	24.44	23.75	0.69	0.15	0.35	4.71

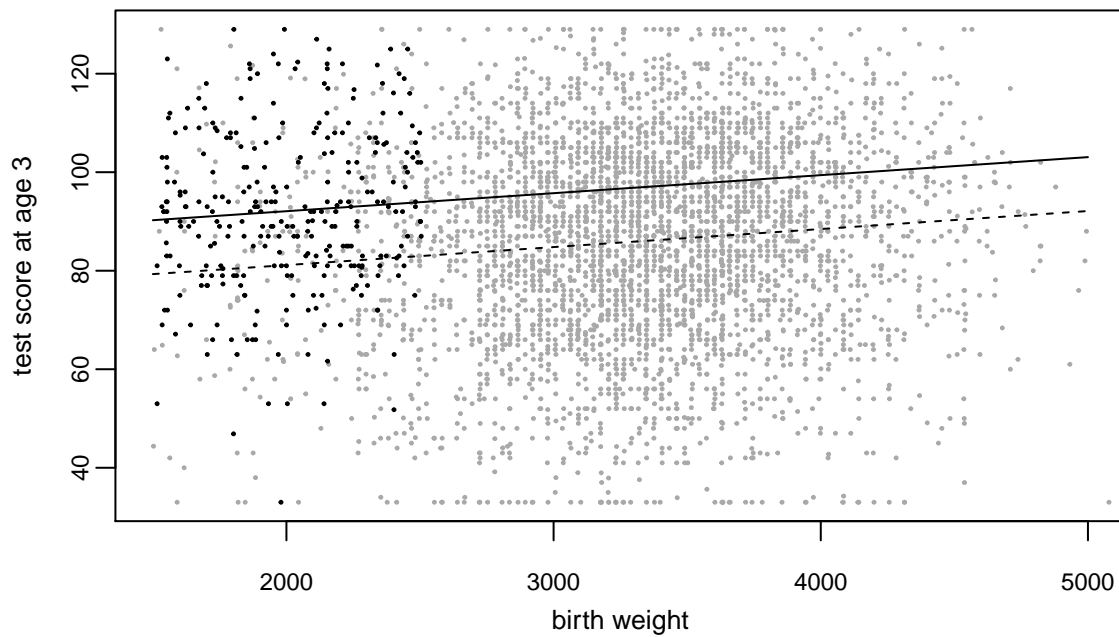
Bivariate

Figure 10.4

matches text

```
# need to jitter things a little

# birthweight
tmp <- lm(ppvtr.36 ~ bw + treat, data = cc2)$coef
# plot(cc2$bw, cc2$ppvtr.36, xlab='birth weight', ylab='test score at age 3',
# cex=.8, mgp=c(2,.5,0), main='', type='n', xlim=c(1500,5000))
plot(cc2$bw, cc2$ppvtr.36, xlab = "birth weight", ylab = "test score at age 3", mgp = c(2,
0.5, 0), main = "", type = "n", xlim = c(1500, 5000), cex.axis = 0.75, cex.lab = 0.8,
lab = c(3, 5, 7), xaxt = "n")
axis(side = 1, at = c(2000, 3000, 4000, 5000), cex.axis = 0.75)
points(cc2$bw[cc2$treat == 0] + runif(sum(cc2$treat == 0), -0.5, 5), cc2$ppvtr.36[cc2$treat ==
0], col = "darkgrey", pch = 20, cex = 0.3)
points(cc2$bw[cc2$treat == 1] + runif(sum(cc2$treat == 1), -0.5, 5), cc2$ppvtr.36[cc2$treat ==
1], pch = 20, cex = 0.3)
# curve(tmp[1]+tmp[2]*x, add=T, col='darkgrey', cex=2)
curve(tmp[1] + tmp[2] * x, add = T, lty = 2)
curve(tmp[1] + tmp[3] + tmp[2] * x, add = T)
```



```
## also tried ineeds, age, momage (shows treats with no comps), momed (excellent
## overlap)
```

```
# income to needs
```

```
tmp <- lm(ppvtr.36 ~ age + treat, data = cc2)$coef
```

```
plot(cc2$age, cc2$ppvtr.36, xlab = "age of child (months) in year 2000", ylab = "test scores at age 3",
     cex = 0.8, mgp = c(2, 0.5, 0), main = "data and estimated regression line", type = "n")
```

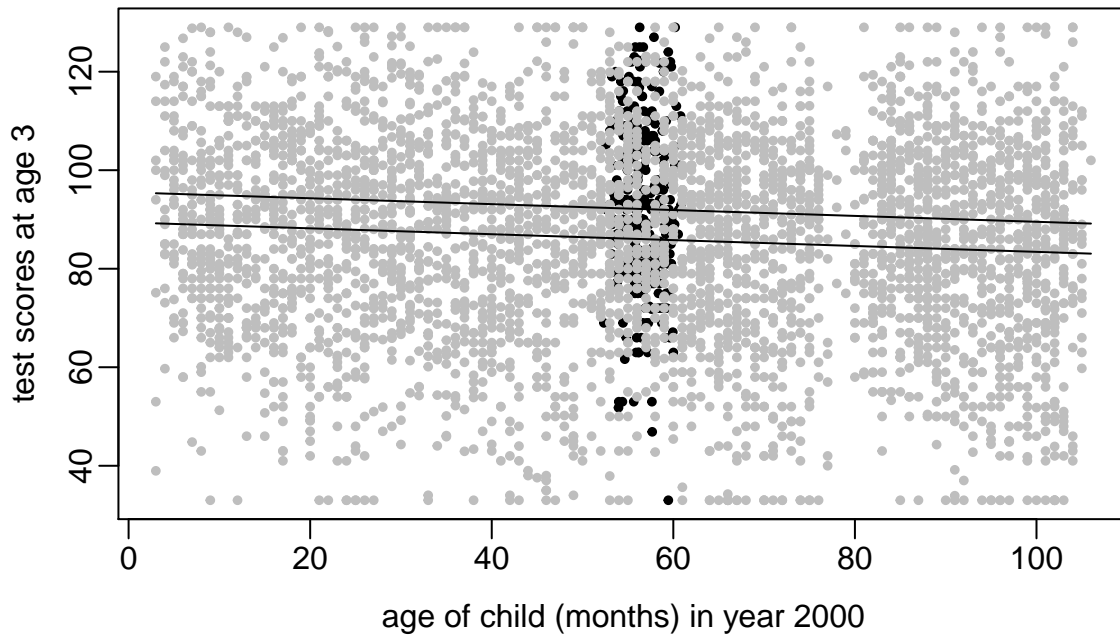
```
points(cc2$age[cc2$treat == 1], cc2$ppvtr.36[cc2$treat == 1], pch = 20, cex = 0.8)
```

```
points(cc2$age[cc2$treat == 0], cc2$ppvtr.36[cc2$treat == 0], cex = 0.8, pch = 20,
      col = "gray")
```

```
curve(tmp[1] + tmp[2] * x, add = T)
```

```
curve(tmp[1] + tmp[3] + tmp[2] * x, add = T)
```

data and estimated regression line



Stratification

Figure 10.6

matches text

```
tes.ns = matrix(0, 4, 4)
tes.ns[1, ] <- c(mean(cc2$ppvtr.36[cc2$treat == 1 & cc2$lths == 1]) - mean(cc2$ppvtr.36[cc2$treat ==
0 & cc2$lths == 1]), sum(cc2$treat == 1 & cc2$lths == 1), sum(cc2$lths == 1),
var(cc2$ppvtr.36[cc2$treat == 1 & cc2$lths == 1])/sum(cc2$treat == 1 & cc2$lths ==
1) + var(cc2$ppvtr.36[cc2$treat == 0 & cc2$lths == 1])/sum(cc2$treat == 0 &
cc2$lths == 1)))
tes.ns[2, ] <- c(mean(cc2$ppvtr.36[cc2$treat == 1 & cc2$hs == 1]) - mean(cc2$ppvtr.36[cc2$treat ==
0 & cc2$hs == 1]), sum(cc2$treat == 1 & cc2$hs == 1), sum(cc2$hs == 1), var(cc2$ppvtr.36[cc2$treat ==
1 & cc2$hs == 1])/sum(cc2$treat == 1 & cc2$hs == 1) + var(cc2$ppvtr.36[cc2$treat ==
0 & cc2$hs == 1])/sum(cc2$treat == 0 & cc2$hs == 1)))
tes.ns[3, ] <- c(mean(cc2$ppvtr.36[cc2$treat == 1 & cc2$ltcoll == 1]) - mean(cc2$ppvtr.36[cc2$treat ==
0 & cc2$ltcoll == 1]), sum(cc2$treat == 1 & cc2$ltcoll == 1), sum(cc2$ltcoll ==
1), var(cc2$ppvtr.36[cc2$treat == 1 & cc2$ltcoll == 1])/sum(cc2$treat == 1 &
cc2$ltcoll == 1) + var(cc2$ppvtr.36[cc2$treat == 0 & cc2$ltcoll == 1])/sum(cc2$treat ==
0 & cc2$ltcoll == 1)))
tes.ns[4, ] <- c(mean(cc2$ppvtr.36[cc2$treat == 1 & cc2$college == 1]) - mean(cc2$ppvtr.36[cc2$treat ==
0 & cc2$college == 1]), sum(cc2$treat == 1 & cc2$college == 1), sum(cc2$college ==
1), var(cc2$ppvtr.36[cc2$treat == 1 & cc2$college == 1])/sum(cc2$treat == 1 &
cc2$college == 1) + var(cc2$ppvtr.36[cc2$treat == 0 & cc2$college == 1])/sum(cc2$treat ==
0 & cc2$college == 1)))
```

```
temp <- tes.ns
rownames(temp) <- c("lths", "hs", "ltcoll", "college")
# calc se from var
temp[, 4] <- sqrt(tes.ns[, 4])
kable(temp, digits = 2, col.names = c("trt_eff", "n_trt", "n_ctrl", "se"))
```

	trt_eff	n_trt	n_ctrl	se
lths	9.30	126	1358	1.34
hs	4.06	82	1820	1.82
ltcoll	7.87	48	837	2.31
college	4.62	34	366	2.14

Formula 10.2

matches text

```
## now overall te.s and vars
c(tes.ns[, 1]) %*% tes.ns[, 2]/sum(tes.ns[, 2])
```

```
##           [,1]
## [1,] 7.032178
```

```
# 7.03
```

```
c(tes.ns[, 1]) %*% tes.ns[, 3]/sum(tes.ns[, 3])
```

```
##           [,1]
## [1,] 6.457971
```

```
# 6.46
```

```
sqrt((c(tes.ns[, 2]^2) %*% (tes.ns[, 4]))/(sum(tes.ns[, 2])^2))
```

```
##           [,1]
## [1,] 0.9016809
```

```
# .90
```

```
# num=(126^2)*(1.34^2) + (82^2)*(1.82^2) + (48^2)*(2.31^2) + (34^2)*(2.14^2)
num = (126^2) * (1.8) + (82^2) * (3.31) + (48^2) * (5.33) + (34^2) * (4.58)
denom = (126 + 82 + 48 + 34)^2
```

```
sqrt(num/denom)
```

```
## [1] 0.9018942
```

```
reg.strat <- lm(ppvtr.36 ~ treat + as.factor(educ), data = cc2)
# 6.99
```

```
### by birthweight group
```

```
tes.ns2 = matrix(0, 2, 3)
```

```
tes.ns2[1, ] <- c(mean(cc2$ppvtr.36[cc2$treat == 1 & cc2$bwg == 1]) - mean(cc2$ppvtr.36[cc2$treat == 0 & cc2$bwg == 1]), sum(cc2$treat == 1 & cc2$bwg == 1), sum(cc2$bwg == 1))
```

```
tes.ns2[2, ] <- c(mean(cc2$ppvtr.36[cc2$treat == 1 & cc2$bwg == 0]) - mean(cc2$ppvtr.36[cc2$treat == 0 & cc2$bwg == 0]), sum(cc2$treat == 1 & cc2$bwg == 0), sum(cc2$bwg == 0))
```



```

round(tes.ns2, 2)

##      [,1] [,2] [,3]
## [1,]  7.64 142 4159
## [2,] 10.87 148  222
# [,1] [,2] [,3] [,1] 7.63621 142 4159 [2,] 10.39012 235 352

c(tes.ns2[, 1]) %>% tes.ns2[, 2]/sum(tes.ns2[, 2])

##      [,1]
## [1,] 9.284831
# 9.28

c(tes.ns2[, 1]) %>% tes.ns2[, 3]/sum(tes.ns2[, 3])

##      [,1]
## [1,] 7.799905
# 7.8

```

P-Score Matchinng

```

covs <- c("bw", "bwg", "hispanic", "black", "white", "b.marr", "lths", "hs", "ltcoll",
  "college", "work.dur", "prenatal", "momage", "sex", "first", "preterm", "age",
  "dayskidh")

##### trying to fit a better model to deal with these issues

### transformed variables now created at top of file

# ps.fit.2 <- glm(treat ~ bwg*as.factor(educ) + as.factor(ethnic)*b.marr +
# work.dur + prenatal + preterm + age + momage + sex + first + bw + dayskidT +
# preterm + pretermT + momage + momageT + black*(bw + preterm +dayskidT) +
# b.marr*(bw + preterm +dayskidT), data=cc2,family=binomial)

ps.fit.2 <- glm(treat ~ work.dur + prenatal + preterm + age + momage + sex + first +
  bwT + dayskidT + preterm + pretermT + momage + as.factor(ethnic) * b.marr + b.marr *
  hs + b.marr * college + b.marr * dayskidT, data = cc2, family = binomial)

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

pscores2 <- ps.fit.2$linear
matches2 <- matching(z = cc2$treat, score = pscores2)
matched2 <- cc2[matches2$matched, ]
ps2.mat <- pscores2[matches2$matched]

sum(pscores2[cc2$treat == 1] > max(pscores2[cc2$treat == 0]))

## [1] 16
# 16

reg.ps <- lm(ppvtr.36 ~ treat + bw + bwg + hispanic + black + b.marr + lths + hs +

```

```

    ltcoll + work.dur + prenatal + momage + sex + first + preterm + age + dayskidh,
    data = matched2)
round(summary(reg.ps)$coef[2, ], digits = 2)

##      Estimate Std. Error      t value    Pr(>|t|)
##      10.51      1.55      6.77      0.00
# 10.3 (1.55)

diff.means.matched2 = matrix(0, length(covs2), 6)
for (i in 1:length(covs2)) {
  diff.means.matched2[i, 1:2] <- c(mean(matched2[matched2$treat == 1, covs2[i]]),
    mean(matched2[matched2$treat == 0, covs2[i]]))
  diff.means.matched2[i, 3] <- diff.means.matched2[i, 1] - diff.means.matched2[i,
    2]
  # note in next we still divide it by the standard deviation from full dataset
  diff.means.matched2[i, 5] <- sqrt(var(matched2[matched2$treat == 1, covs2[i]])/sum(cc2$treat ==
    1) + var(cc2[cc2$treat == 0, covs2[i]])/sum(cc2$treat == 0))
  diff.means.matched2[i, 6] <- sqrt((var(cc2[cc2$treat == 1, covs2[i]]) + var(cc2[cc2$treat ==
    0, covs2[i]]))/2)
  diff.means.matched2[i, 4] <- diff.means.matched2[i, 3]/diff.means.matched2[i,
    6]
}
dimnames(diff.means.matched2) <- list(covs2, c("treat", "control", "diff", "diff.std",
  "se", "sd"))
round(diff.means.matched2, 2)

##      treat control  diff diff.std  se  sd
## neg.bw      491.35  233.16 258.19    0.58 18.81 444.70
## preterm      6.07   4.97  1.10    0.56 0.12  1.97
## dayskidh     14.69  10.35  4.34    0.49 0.67  8.86
## sex          0.51   0.53 -0.03   -0.06 0.03  0.50
## first        0.48   0.48  0.01    0.01 0.03  0.50
## age          56.67  56.81 -0.14   -0.01 0.47 20.59
## black        0.50   0.41  0.09    0.20 0.03  0.48
## hispanic     0.09   0.13 -0.04   -0.12 0.02  0.36
## white        0.40   0.46 -0.05   -0.10 0.03  0.50
## b.unmarr     0.57   0.46  0.11    0.22 0.03  0.48
## lths         0.43   0.37  0.07    0.14 0.03  0.48
## hs           0.28   0.38 -0.10   -0.20 0.03  0.47
## ltcoll       0.17   0.19 -0.03   -0.07 0.02  0.38
## college      0.12   0.06  0.06    0.18 0.02  0.30
## work.dur     0.59   0.58  0.01    0.02 0.03  0.49
## no.prenatal  0.04   0.02  0.03    0.17 0.01  0.17
## momage       24.44  23.58  0.86    0.18 0.35  4.71

covs2C <- c("neg.bw", "preterm", "dayskidh", "age", "momage")

ratio.sds2 = matrix(0, length(covs2C), 2)
for (i in 1:length(covs2C)) {
  ratio.sds2[i, 1] <- c(sd(cc2[cc2$treat == 1, covs2C[i]])/sd(cc2[cc2$treat ==
    0, covs2C[i]]))
  ratio.sds2[i, 2] <- c(sd(matched2[matched2$treat == 1, covs2C[i]])/sd(matched2[matched2$treat ==
    0, covs2C[i]]))
}

```

```

dimnames(ratio.sds2) <- list(covs2C, c("unmatched ratio", "matched ratio"))
round(ratio.sds2, 2)

##          unmatched ratio matched ratio
## neg.bw          0.50          0.85
## preterm          0.95          0.85
## dayskidh         2.07          0.88
## age              0.07          0.07
## momage           1.86          1.75
# but we won't use this in the paper -- will use the version in which the pscores
# also had the state indicators
reg.st.ps <- lm(ppvtr.36 ~ treat + hispanic + black + b.marr + lths + hs + ltcoll +
  work.dur + prenatal + momage + sex + first + preterm + age + dayskidh + bw +
  unemp.rt + st5 + st9 + st12 + st25 + st36 + st42 + st48 + st53, data = matched2)
round(summary(reg.st.ps)$coef[2, ], digits = 2)

##      Estimate Std. Error    t value   Pr(>|t|)
##      10.34      2.26        4.58      0.00
# 10.3 (2.26)

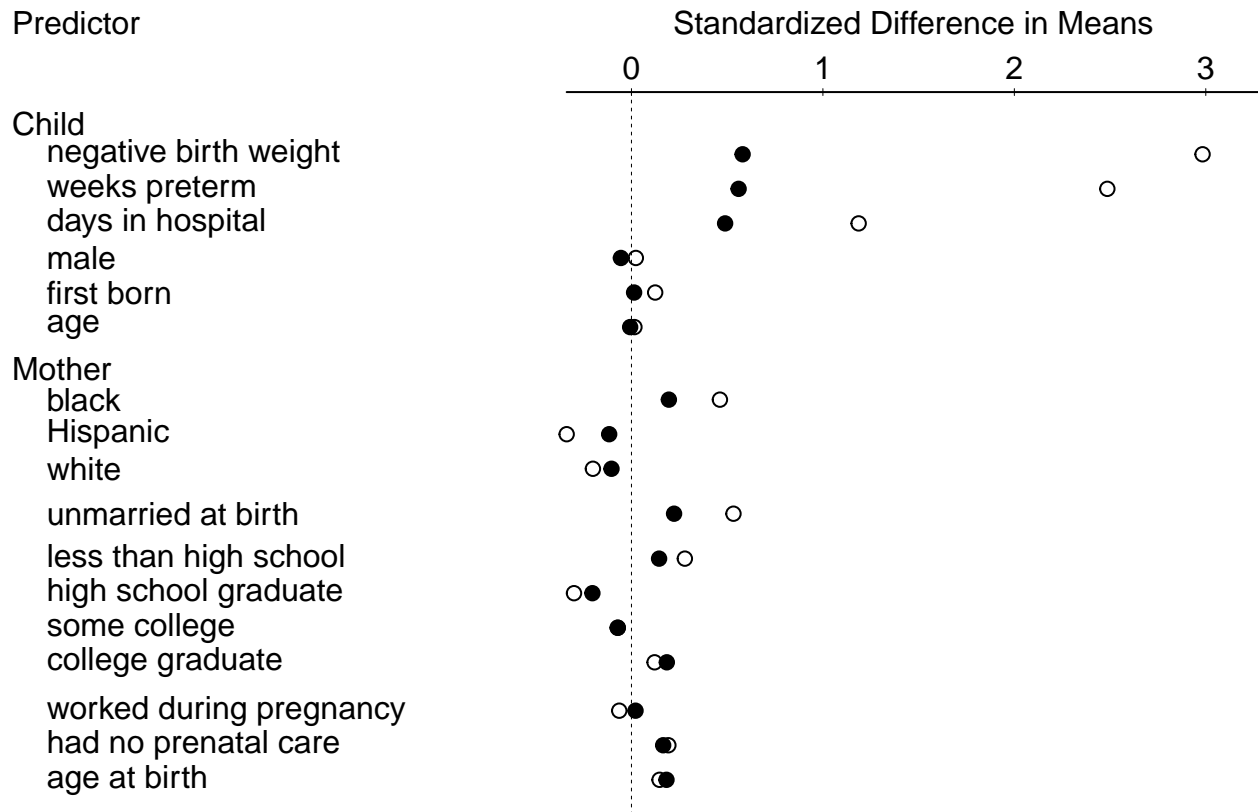
name = cov.nms
est = diff.means[, 3]
sd = diff.means[, 6]
est2 = diff.means.matched2[, 3]
sd2 = diff.means.matched2[, 6]

J <- 20
# data.locations = c(1.8,2.8,3.8,4.8,5.8,6.8,9:19)
data.locations = c(1.8, 2.8, 3.8, 4.8, 5.8, 6.8, 8.9, 9.9, 10.9, 12.2, 13.5, 14.5,
  15.5, 16.5, 17.9, 18.9, 19.9)
name.range <- 0.8
x.range <- range(c(est, est2)/c(sd, sd2))
x.range[2] <- x.range[2] + 0.3
A <- -x.range[1]/(x.range[2] - x.range[1])
B <- 1/(x.range[2] - x.range[1])
height <- 0.35 * J
width <- 3 * (name.range + 1)
#

par(mar = c(0, 0, 0, 0))
plot(c(-name.range, 1), c(3, -J - 2), bty = "n", xlab = "", ylab = "", xaxt = "n",
  yaxt = "n", xaxs = "i", yaxs = "i", type = "n")
text(-name.range, 2, "Predictor", adj = 0, cex = 1)
text(0.5, 2, "Standardized Difference in Means", adj = 0.5, cex = 1)
lines(c(0, 1), c(0, 0))
lines(c(A, A), c(0, -J - 1), lty = 2, lwd = 0.5)
ax <- pretty(x.range)
ax <- ax[(A + B * ax) > 0 & (A + B * ax) < 1]
segments(A + B * ax, -0.1, A + B * ax, 0.1, lwd = 0.5)
text(A + B * ax, 0.7, ax, cex = 1)
text(-name.range, -0.9, "Child", adj = 0, cex = 1)
text(-name.range, -8, "Mother", adj = 0, cex = 1)
text(-name.range + 0.05, -data.locations, name, adj = 0, cex = 1)

```

```
points(A + B * (est/sd), -data.locations)
points(A + B * (est2/sd2), -data.locations, pch = 20, cex = 1.5)
```



```
# if (bottom){ lines (c(0,1), c(-J-1,-J-1)) segments (A + B*ax, -J-1-.1, A +
# B*ax, -J-1+.1, lwd=.5) text (A + B*ax, -J-1-.7, ax, cex=1) }
dev.off()
```

```
## null device
## 1
```

Figure 10.7

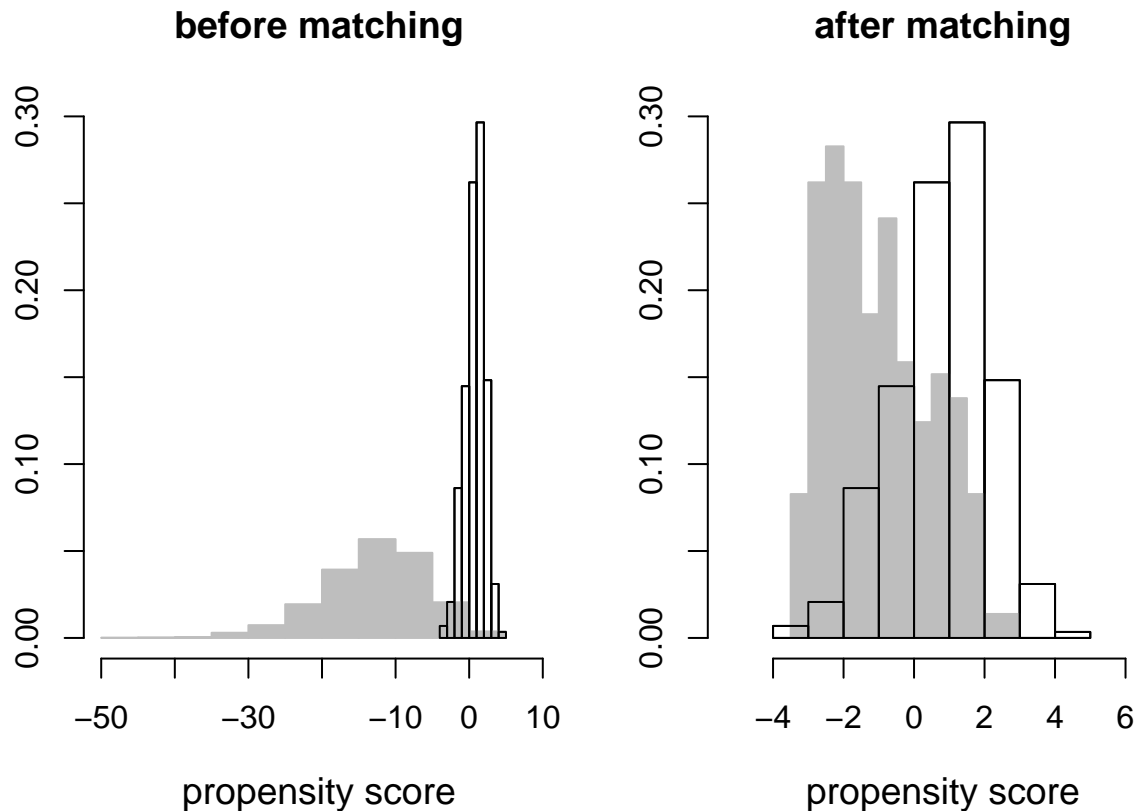
does not match text

```
##### plot of overlap in pscores
```

```
##### THIS NEXT IN PAPER (AFTER CONVERTING TO PDF AND CROPPING)
```

```
par(mfrow = c(1, 2))
par(mar = c(4, 2, 4, 2))
hist(pscores2[cc2$treat == 1], xlim = c(-50, max(pscores2) + 5), main = "before matching",
     xlab = "propensity score", freq = F, cex.main = 1.2, cex.lab = 1.2)
hist(pscores2[cc2$treat == 0 & pscores2 > (-50)], main = "", col = "grey", border = "grey",
     freq = F, add = T)
hist(pscores2[cc2$treat == 1], xlim = c(-50, max(pscores2) + 5), main = "before matching",
     xlab = "propensity score", freq = F, cex.main = 1.2, cex.lab = 1.2, add = TRUE)
hist(ps2.mat[matched2$treat == 1], xlim = c(min(ps2.mat) - 2, 7), main = "after matching",
     xlab = "propensity score", freq = F, cex.main = 1.2, cex.lab = 1.2)
```

```
# hist(ps2.mat[matched2$treat==1],xlim=c(-40,7),main='after matching',
# xlab='propensity score',freq=F,cex.main=1.2,cex.lab=1.2)
hist(ps2.mat[matched2$treat == 0], main = "", freq = F, add = T, col = "grey", border = "grey")
hist(ps2.mat[matched2$treat == 1], freq = F, cex.main = 1.2, cex.lab = 1.2, add = TRUE)
```



```
dev.off()

## null device
##      1

sum(pscores2 < (-50))

## [1] 6

# 6 controls deleted from the plot to aid in seeing differences across groups
# plot of overlap in pcores FREQUENCY NOT DENSITY

par(mfrow = c(1, 2))
par(mar = c(4, 2, 4, 2))
hist(pscores2[cc2$treat == 0], xlim = c(min(pscores2), max(pscores2)), main = "before matching",
     xlab = "propensity score", nclass = 20, border = "darkgrey", cex.main = 1.2,
     cex.lab = 1.2, col = "darkgrey")
hist(pscores2[cc2$treat == 1], add = T)
hist(ps2.mat[matched2$treat == 0], xlim = c(min(ps2.mat) - 2, max(pscores2)), main = "after matching",
     xlab = "propensity score", nclass = 10, border = "grey", cex.main = 1.2, cex.lab = 1.2,
     col = "grey")
hist(ps2.mat[matched2$treat == 1], nclass = 20, add = T)
dev.off()
```

```
## null device
##          1

pscores2I = invlogit(pscores2)
par(mfrow = c(1, 2))
par(mar = c(4, 2, 4, 2))
hist(pscores2I[cc2$treat == 0 & pscores2I > 0.01], xlim = c(0.01, 1), main = "before matching",
     xlab = "propensity score", nclass = 50, border = "darkgrey", cex.main = 1.2,
     cex.lab = 1.2, col = "darkgrey")
hist(pscores2I[cc2$treat == 1], add = T, nclass = 50)
hist(invlogit(ps2.mat[matched2$treat == 0]), xlim = c(0.01, 1), main = "after matching",
     xlab = "propensity score", nclass = 50, border = "darkgrey", cex.main = 1.2,
     cex.lab = 1.2, col = "darkgrey")
hist(invlogit(ps2.mat[matched2$treat == 1]), nclass = 50, add = T)
dev.off()

## null device
##          1

sum(pscores2I < 0.01)

## [1] 3652
# removing the 3652 observations from the plot with pscores<.01

sum(pscores2[cc2$treat == 1] > max(pscores2[cc2$treat == 0]))

## [1] 16
```

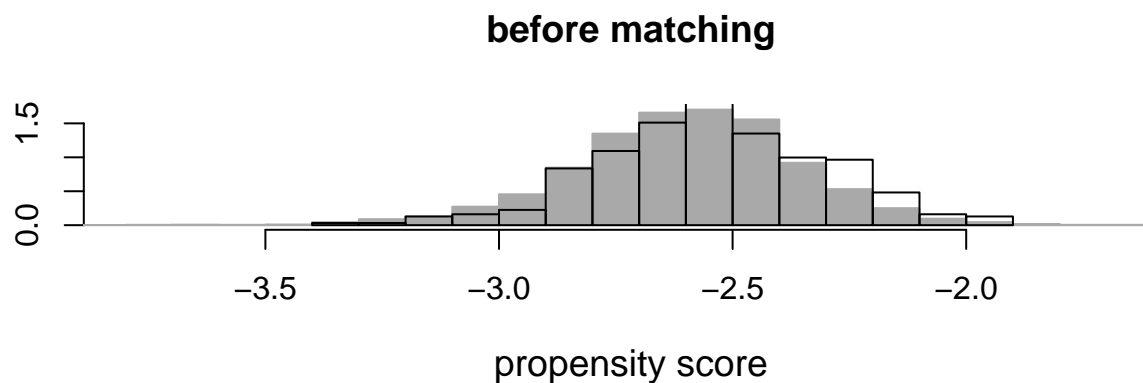
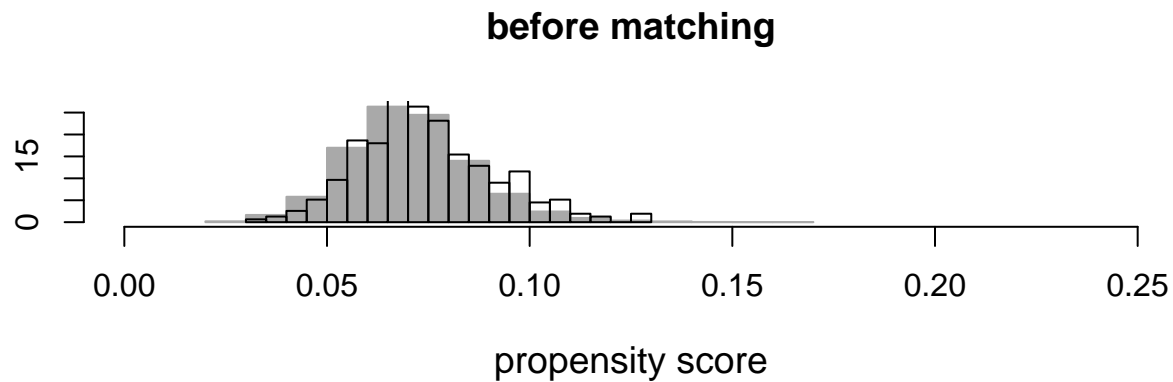
UNIVARIATE

```
### overfit diagnostic
set.seed(1234)
cc2.refit = cbind.data.frame(cc2, treat.rand = rbinom(nrow(cc2), 1, mean(cc2$treat)))
ps.refit.2 <- glm(treat.rand ~ work.dur + prenatal + preterm + age + momage + sex +
  first + bwT + dayskidT + preterm + pretermT + momage + as.factor(ethnic) * b.marr +
  b.marr * hs + b.marr * college + b.marr * dayskidT, data = cc2.refit, family = binomial)
# ps.refit.2 <- glm(treat ~ work.dur + prenatal + preterm + age + momage + sex +
# first + bwT + dayskidT + preterm + pretermT + momage + as.factor(ethnic)*b.marr
# + b.marr*hs + b.marr*college + b.marr*dayskidT, data=cc2.refit,family=binomial)

pscores2.rf = ps.refit.2$linear
pscores2I.rf = invlogit(pscores2.rf)

par(mfrow = c(2, 1))
par(mar = c(4, 2, 4, 2))
hist(pscores2I.rf[cc2.refit$treat.rand == 0], xlim = c(0, 0.25), main = "before matching",
     xlab = "propensity score", nclass = 20, border = "darkgrey", cex.main = 1.2,
     cex.lab = 1.2, col = "darkgrey", freq = FALSE)
hist(pscores2I.rf[cc2.refit$treat.rand == 1], add = T, nclass = 20, freq = FALSE)
hist(pscores2I.rf[cc2.refit$treat.rand == 0], xlim = c(min(pscores2.rf), max(pscores2.rf)),
     main = "before matching", xlab = "propensity score", nclass = 20, border = "darkgrey",
```

```
cex.main = 1.2, cex.lab = 1.2, col = "darkgrey", freq = FALSE)
hist(pscores2.rf[cc2.refit$treat.rand == 1], add = T, nclass = 20, freq = FALSE)
```



```
# hist(invlogit(ps2.mat[matched2$treat==0]),xlim=c(0.03,1),main='after
# matching',xlab='propensity score',
# nclass=100,border='darkgrey',cex.main=1.2,cex.lab=1.2,col='darkgrey')
# hist(invlogit(ps2.mat[matched2$treat==1]),nclass=100,add=T)

hist(pscores2.rf[cc2.refit$treat.rand == 0], xlim = c(-4, -1.5), ylim = c(0, 2),
     main = "before matching", xlab = "propensity score", nclass = 20, border = "darkgrey",
     cex.main = 1.2, cex.lab = 1.2, col = "darkgrey", freq = FALSE)
hist(pscores2.rf[cc2.refit$treat.rand == 1], add = T, nclass = 20, freq = FALSE)

# evaluate wrt the randomization distribution
sum(pscores2.rf[cc2.refit$treat.rand == 1] > max(pscores2.rf[cc2.refit$treat.rand ==
0]))

## [1] 0

nitters = 1000
not.comm = rep(0, nitters)
for (i in 1:nitters) {
  cc2.refit.i = cbind.data.frame(cc2, treat.rand = rbinom(nrow(cc2), 1, mean(cc2$treat)))
  ps.refit.2.i <- glm(treat.rand ~ work.dur + prenatal + preterm + age + momage +
    sex + first + bwT + dayskidT + preterm + pretermT + momage + as.factor(ethnic) *
    b.marr + b.marr * hs + b.marr * college + b.marr * dayskidT, data = cc2.refit.i,
    family = binomial)
```

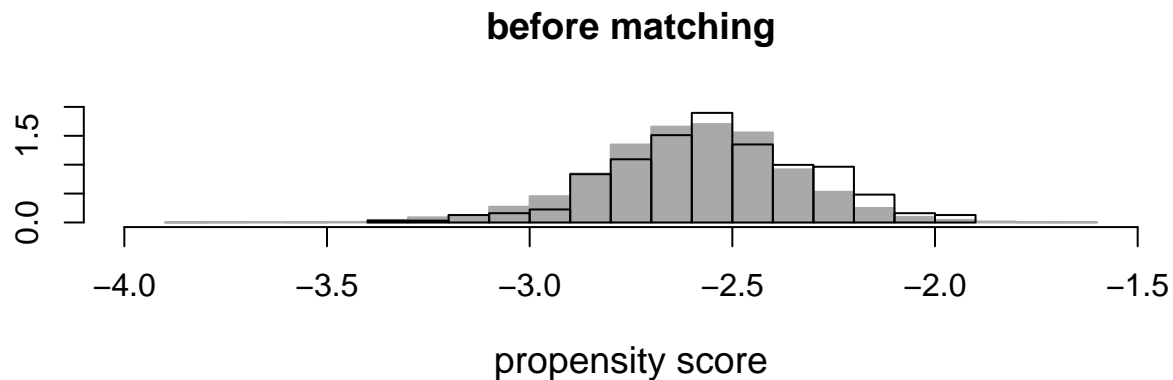
```

pscores2.rf.i = ps.refit.2.i$linear
not.comm[i] = sum(pscores2.rf.i[cc2.refit.i$treat.rand == 1] > max(pscores2.rf.i[cc2.refit.i$treat.
0]))
}

```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```



Geographic information

does not match text, slightly greater (8.9 vs 8.8)

```

##### and a third version to deal with the geographic location DOESN'T WORK WELL --
##### WON'T USE

```

```

ps.fit.3 <- glm(treat ~ work.dur + prenatal + preterm + age + momage + sex + first +
  bwT + dayskidT + preterm + pretermT + momage + as.factor(ethnic) * b.marr + b.marr *
  hs + b.marr * college + b.marr * dayskidT + st5 + st9 + st12 + st25 + st36 +
  st42 + st48 + st53, data = cc2, family = binomial)

```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```

# ps.fit.3 <- glm(treat ~ work.dur + prenatal + preterm + age + momage + sex +
# first + bwT + dayskidT + preterm + pretermT + momage + as.factor(ethnic)*b.marr
# + b.marr*hs + b.marr*college + b.marr*dayskidT + st5 + st9 + st12 + st25 + st36
# + st42 + st48 + st53, data=cc2,family=binomial) 10.22

```



```

pscores3 <- ps.fit.3$linear
matches3 <- matching(z = cc2$treat, score = pscores3)
matched3 <- cc2[matches3$matched, ]
ps3.mat <- pscores3[matches3$matched]

reg.st.ps <- lm(ppvtr.36 ~ treat + hispanic + black + b.marr + lths + hs + ltcoll +
  work.dur + prenatal + momage + sex + first + preterm + age + dayskidh + bw +
  st5 + st9 + st12 + st25 + st36 + st42 + st48 + st53, data = matched3)
round(summary(reg.st.ps)$coef[2, ], digits = 2)

##      Estimate Std. Error      t value    Pr(>|t|)
##          8.92       2.09         4.27      0.00
# 8.92 (2.09)

# umemp.rt?
diff.means.matched3 = matrix(0, length(covs2), 6)
for (i in 1:length(covs2)) {
  diff.means.matched3[i, 1:2] <- c(mean(matched3[matched3$treat == 1, covs2[i]]),
    mean(matched3[matched3$treat == 0, covs2[i]]))
  diff.means.matched3[i, 3] <- diff.means.matched3[i, 1] - diff.means.matched3[i,
    2]
  # note in next we still divide it by the standard deviation from full dataset
  diff.means.matched3[i, 5] <- sqrt(var(matched3[matched3$treat == 1, covs2[i]])/sum(cc2$treat ==
    1) + var(cc2[cc2$treat == 0, covs2[i]]/sum(cc2$treat == 0))
  diff.means.matched3[i, 6] <- sqrt((var(cc2[cc2$treat == 1, covs2[i]]) + var(cc2[cc2$treat ==
    0, covs2[i]]))/2)
  diff.means.matched3[i, 4] <- diff.means.matched3[i, 3]/diff.means.matched3[i,
    6]
}
dimnames(diff.means.matched3) <- list(covs2, c("treat", "control", "diff", "diff.std",
  "se", "sd"))
round(diff.means.matched3, 2)

##          treat control  diff diff.std  se  sd
## neg.bw      491.35 -167.57 658.92    1.48 18.81 444.70
## preterm        6.07   2.90  3.17    1.61  0.12  1.97
## dayskidh     14.69   6.79  7.90    0.89  0.67  8.86
## sex           0.51   0.54 -0.03   -0.06  0.03  0.50
## first        0.48   0.44  0.04    0.09  0.03  0.50
## age         56.67  56.99 -0.32   -0.02  0.47 20.59
## black        0.50   0.40  0.10    0.22  0.03  0.48
## hispanic     0.09   0.24 -0.14   -0.41  0.02  0.36
## white        0.40   0.36  0.04    0.08  0.03  0.50
## b.unmarr     0.57   0.43  0.13    0.28  0.03  0.48
## lths         0.43   0.36  0.08    0.17  0.03  0.48
## hs           0.28   0.36 -0.07   -0.15  0.03  0.47
## ltcoll       0.17   0.23 -0.06   -0.16  0.02  0.38
## college      0.12   0.06  0.06    0.18  0.02  0.30
## work.dur     0.59   0.57  0.02    0.04  0.03  0.49
## no.prenatal  0.04   0.02  0.03    0.17  0.01  0.17
## momage      24.44  23.69  0.76    0.16  0.35  4.71

ratio.sds3 = matrix(0, length(covs2), 2)
for (i in 1:length(covs2)) {

```

```

ratio.sds3[i, 1] <- c(sd(cc2[cc2$treat == 1, covs2[i]])/sd(cc2[cc2$treat == 0,
covs2[i]]))
ratio.sds3[i, 2] <- c(sd(matched3[matched3$treat == 1, covs2[i]])/sd(matched3[matched3$treat ==
0, covs2[i]]))
}
dimnames(ratio.sds3) <- list(covs2, c("unmatched ratio", "matched ratio"))
round(ratio.sds3, 2)

```

```

##          unmatched ratio matched ratio
## neg.bw          0.50          0.72
## preterm          0.95          0.86
## dayskidh         2.07          0.96
## sex              1.00          1.00
## first            1.01          1.01
## age              0.07          0.07
## black            1.11          1.02
## hispanic         0.71          0.68
## white            0.98          1.02
## b.unmarr         1.07          1.00
## lths             1.08          1.04
## hs               0.91          0.94
## ltcoll           0.94          0.89
## college          1.18          1.33
## work.dur         1.02          0.99
## no.prenatal      1.85          1.59
## momage           1.86          1.79

```