RDD2

Calculate a propensity score using OLS using up to a quadratic

```
read_data <- function(df)
{
  full_path <- paste("https://raw.github.com/scunning1975/mixtape/master/",
                     df, sep = "")
  df <- read_dta(full_path)
  return(df)
}

nsw_dw <- read_data("nsw_mixtape.dta")

nsw_dw %>%
  filter(treat == 1) %>%
  summary(re78)
```

```
##    data_id              treat        age             educ
##  Length:185         Min.   :1   Min.   :17.00   Min.   : 4.00
##  Class :character   1st Qu.:1   1st Qu.:20.00   1st Qu.: 9.00
##  Mode  :character   Median :1   Median :25.00   Median :11.00
##                     Mean   :1   Mean   :25.82   Mean   :10.35
##                     3rd Qu.:1   3rd Qu.:29.00   3rd Qu.:12.00
##                     Max.   :1   Max.   :48.00   Max.   :16.00
##      black             hisp              marr           nodegree
##  Min.   :0.0000   Min.   :0.00000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:1.0000   1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :1.0000   Median :0.00000   Median :0.0000   Median :1.0000
##  Mean   :0.8432   Mean   :0.05946   Mean   :0.1892   Mean   :0.7081
##  3rd Qu.:1.0000   3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.00000   Max.   :1.0000   Max.   :1.0000
##       re74            re75            re78
##  Min.   :    0   Min.   :    0   Min.   :    0.0
##  1st Qu.:    0   1st Qu.:    0   1st Qu.:  485.2
##  Median :    0   Median :    0   Median : 4232.3
##  Mean   : 2096   Mean   : 1532   Mean   : 6349.1
##  3rd Qu.: 1291   3rd Qu.: 1817   3rd Qu.: 9643.0
##  Max.   :35040   Max.   :25142   Max.   :60307.9
```

```
mean1 <- nsw_dw %>%
  filter(treat == 1) %>%
  pull(re78) %>%
  mean()

nsw_dw$y1 <- mean1
```

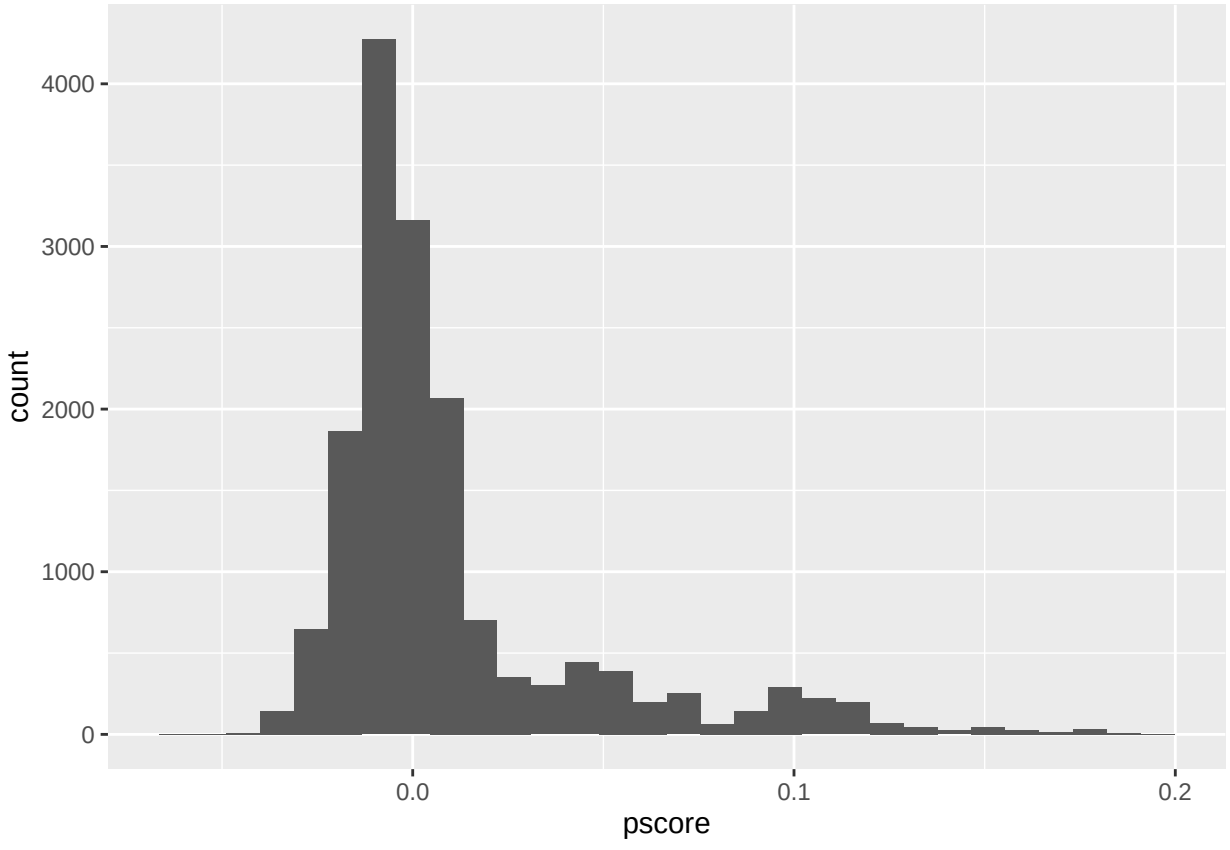
```
nsw_dw %>%
  filter(treat == 0) %>%
  summary(re78)
```

```
##    data_id              treat        age              educ
##  Length:260         Min.   :0   Min.   :17.00   Min.   : 3.00
##  Class :character   1st Qu.:0   1st Qu.:19.00   1st Qu.: 9.00
##  Mode  :character   Median :0   Median :24.00   Median :10.00
##                     Mean   :0   Mean   :25.05   Mean   :10.09
##                     3rd Qu.:0   3rd Qu.:28.00   3rd Qu.:11.00
##                     Max.   :0   Max.   :55.00   Max.   :14.00
##      black             hisp             marr           nodegree
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:1.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:1.0000
##  Median :1.0000   Median :0.0000   Median :0.0000   Median :1.0000
##  Mean   :0.8269   Mean   :0.1077   Mean   :0.1538   Mean   :0.8346
##  3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##       re74              re75              re78             y1
##  Min.   :    0.0   Min.   :    0.0   Min.   :    0   Min.   :6349
##  1st Qu.:    0.0   1st Qu.:    0.0   1st Qu.:    0   1st Qu.:6349
##  Median :    0.0   Median :    0.0   Median : 3139   Median :6349
##  Mean   : 2107.0   Mean   : 1266.9   Mean   : 4555   Mean   :6349
##  3rd Qu.:  139.4   3rd Qu.:  650.1   3rd Qu.: 7288   3rd Qu.:6349
##  Max.   :39570.7   Max.   :23032.0   Max.   :39484   Max.   :6349
```

```
mean0 <- nsw_dw %>%
  filter(treat == 0) %>%
  pull(re78) %>%
  mean()

nsw_dw$y0 <- mean0

ate <- unique(nsw_dw$y1 - nsw_dw$y0) # Before and after first difference for each unit?

nsw_dw <- nsw_dw %>%
  filter(treat == 1) %>%
  select(-y1, -y0)

nsw_dw_cpscontrol <- read_data("cps_mixtape.dta") %>%
  bind_rows(nsw_dw) %>%
  mutate(agesq = age^2,
         agecube = age^3,
         educsq = educ*educ,
         u74 = case_when(re74 == 0 ~ 1, TRUE ~ 0),
         u75 = case_when(re75 == 0 ~ 1, TRUE ~ 0),
         interaction1 = educ*re74,
         re74sq = re74^2,
         re75sq = re75^2,
         interaction2 = u74*hisp)
```

```
ols_nsw_quad <- lm(treat ~ age + agesq + educ + educsq + marr + nodegree + black + hisp + re74 + re75 +

nsw_dw_cpscontrol <- nsw_dw_cpscontrol %>%
  mutate(pscore = ols_nsw_quad$fitted.values)

pscore_control <- nsw_dw_cpscontrol %>%
  filter(treat == 0) %>%
  pull(pscore) %>%
  mean()

pscore_treated <- nsw_dw_cpscontrol %>%
  filter(treat == 1) %>%
  pull(pscore) %>%
  mean()

nsw_dw_cpscontrol %>%
  filter(treat == 0) %>%
  ggplot() +
  geom_histogram(aes(x = pscore))
```

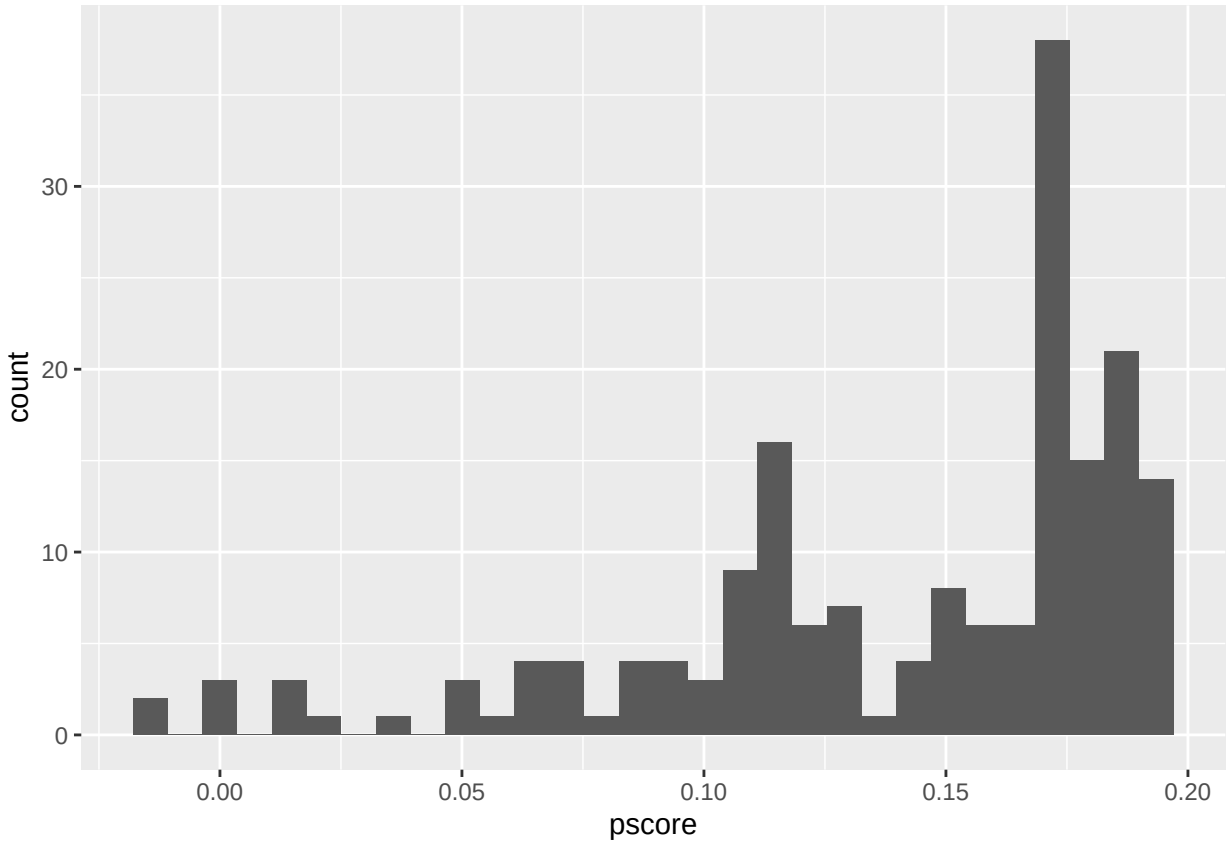
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
nsw_dw_cpscontrol %>%
  filter(treat == 1) %>%
```

```
  ggplot() +
  geom_histogram(aes(x = pscore))
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
### Max and Min values of the scores for treated and control
nsw_dw_cpscontroltreated <- nsw_dw_cpscontrol %>%
  filter(treat == 1)

max(nsw_dw_cpscontroltreated$pscore)
```

```
## [1] 0.1938788
```

```
min(nsw_dw_cpscontroltreated$pscore)
```

```
## [1] -0.01385616
```

```
nsw_dw_cpscontroluntreated<- nsw_dw_cpscontrol %>%
  filter(treat == 0)

max(nsw_dw_cpscontroluntreated$pscore)
```

```
## [1] 0.1938197
```

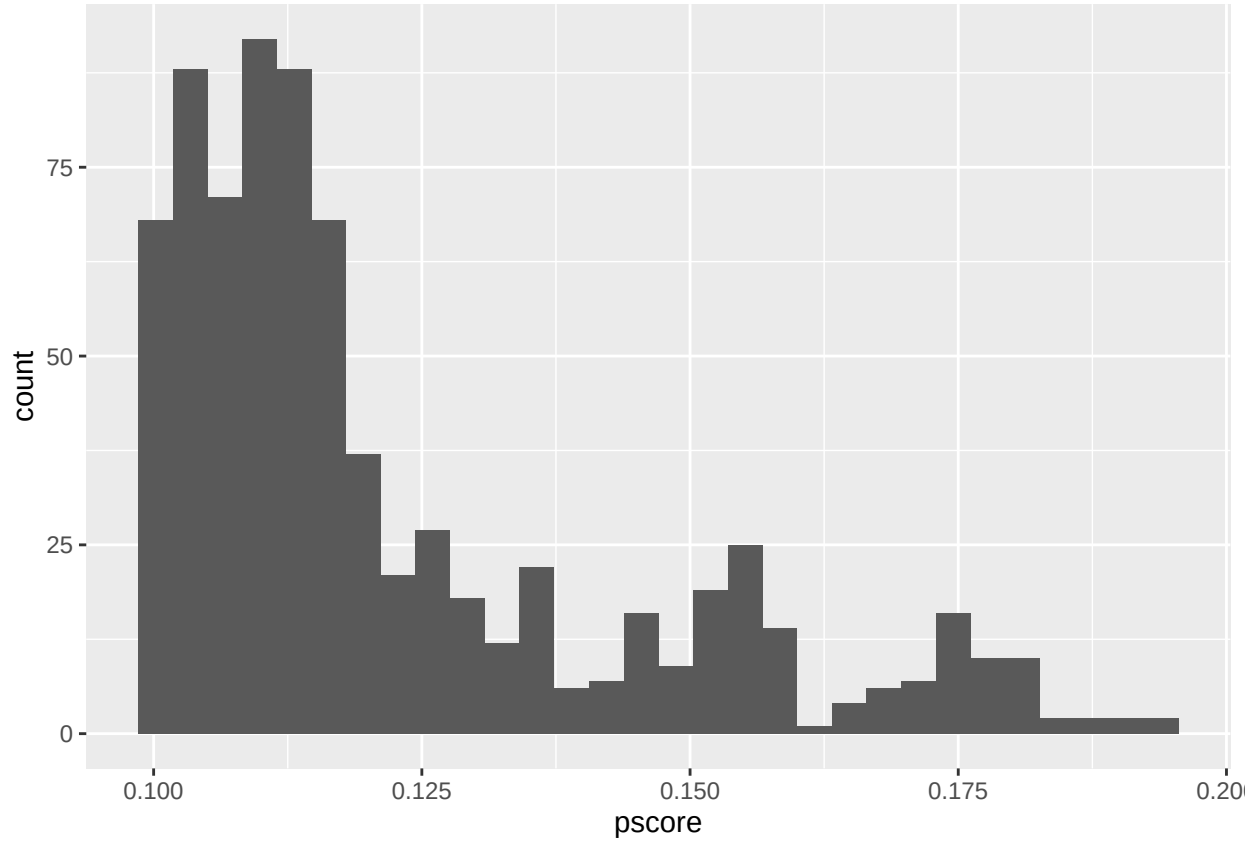
```
min(nsw_dw_cpscontroluntreated$pscore)
```

```
## [1] -0.06371947
```

Droping propensity score

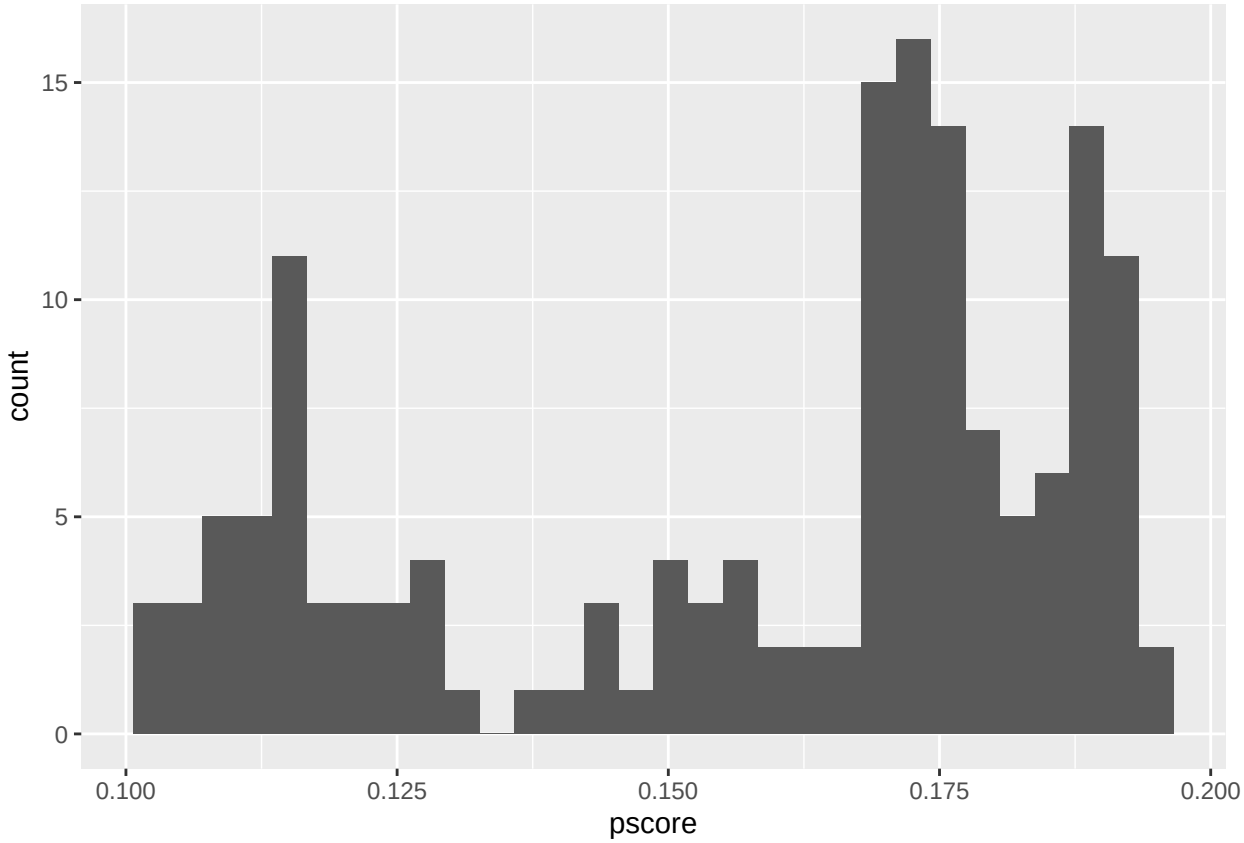
```
nsw_dw_cpscontrolcut <- nsw_dw_cpscontrol %>%
  filter(!(pscore >= 0.9)) %>%
  filter(!(pscore <= 0.1))

nsw_dw_cpscontrolcut %>%
  filter(treat == 0) %>%
  ggplot() +
  geom_histogram(aes(x = pscore))
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
nsw_dw_cpscontrolcut %>%
  filter(treat == 1) %>%
  ggplot() +
  geom_histogram(aes(x = pscore))
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
nsw_dw_cpscontroltreatedcut <- nsw_dw_cpscontrolcut %>%
  filter(treat == 1)

max(nsw_dw_cpscontroltreatedcut$pscore)
```

```
## [1] 0.1938788
```

```
min(nsw_dw_cpscontroltreatedcut$pscore)
```

```
## [1] 0.1011927
```

```
nsw_dw_cpscontroluntreatedcut <- nsw_dw_cpscontrolcut %>%
  filter(treat == 0)

max(nsw_dw_cpscontroluntreatedcut$pscore)
```

```
## [1] 0.1938197
```

```
min(nsw_dw_cpscontroluntreatedcut$pscore)
```

```
## [1] 0.1001003
```

```r
N <- nrow(nsw_dw_cpscontrol)

#- Manual with non-normalized weights using all data
nsw_dw_cpscontrol <- nsw_dw_cpscontrol %>%
  mutate(d1 = treat/pscore,
         d0 = (1-treat)/(1-pscore))
s1 <- sum(nsw_dw_cpscontrol$d1)
s0 <- sum(nsw_dw_cpscontrol$d0)


nsw_dw_cpscontrol <- nsw_dw_cpscontrol %>%
  mutate(y1 = treat * re78/pscore,
         y0 = (1-treat) * re78/(1-pscore),
         ht = y1 - y0)

#- Manual with normalized weights
nsw_dw_cpscontrol <- nsw_dw_cpscontrol %>%
  mutate(y1 = (treat*re78/pscore)/(s1/N),
         y0 = ((1-treat)*re78/(1-pscore))/(s0/N),
         norm = y1 - y0)

nsw_dw_cpscontrol %>%
  pull(ht) %>%
  mean()
```

```
## [1] -15066.65
```

```r
nsw_dw_cpscontrol %>%
  pull(norm) %>%
  mean()
```

```
## [1] -24136.16
```

```r
nsw_dw_cpscontrol %>% pull(norm) %>% mean() - nsw_dw_cpscontrol %>% pull(ht) %>% mean()
```

```
## [1] -9069.51
```

```r
# Trimming data
nsw_dw_cpscontrol <- nsw_dw_cpscontrol %>%
  select(-d1, -d0, -y1, -y0, -ht, -norm) %>%
  filter(!(pscore >= 0.9)) %>%
  filter(!(pscore <= 0.1))

N <- nrow(nsw_dw_cpscontrol)

#- Manual with non-normalized weights using trimmed data
nsw_dw_cpscontrol <- nsw_dw_cpscontrol %>%
  mutate(d1 = treat/pscore,
         d0 = (1-treat)/(1-pscore))

s1 <- sum(nsw_dw_cpscontrol$d1)
```

```
s0 <- sum(nsw_dw_cpscontrol$d0)

nsw_dw_cpscontrol <- nsw_dw_cpscontrol %>%
  mutate(y1 = treat * re78/pscore,
         y0 = (1-treat) * re78/(1-pscore),
         ht = y1 - y0)

#- Manual with normalized weights with trimmed data
nsw_dw_cpscontrol <- nsw_dw_cpscontrol %>%
  mutate(y1 = (treat*re78/pscore)/(s1/N),
         y0 = ((1-treat)*re78/(1-pscore))/(s0/N),
         norm = y1 - y0)

nsw_dw_cpscontrol %>%
  pull(ht) %>%
  mean()
```

```
## [1] -3455.285
```

```
nsw_dw_cpscontrol %>%
  pull(norm) %>%
  mean()
```

```
## [1] -4577.043
```

```
nsw_dw_cpscontrol %>% pull(norm) %>% mean() - nsw_dw_cpscontrol %>% pull(ht) %>% mean()
```

```
## [1] -1121.757
```

Logit Up To Cube

```
read_data <- function(df)
{
  full_path <- paste("https://raw.github.com/scunning1975/mixtape/master/",
                     df, sep = "")
  df <- read_dta(full_path)
  return(df)
}

nsw_dw <- read_data("nsw_mixtape.dta")

nsw_dw %>%
  filter(treat == 1) %>%
  summary(re78)
```

```
##    data_id              treat        age              educ
##  Length:185         Min.   :1   Min.   :17.00   Min.   : 4.00
##  Class :character   1st Qu.:1   1st Qu.:20.00   1st Qu.: 9.00
##  Mode  :character   Median :1   Median :25.00   Median :11.00
```

```
##                    Mean   :1   Mean   :25.82   Mean   :10.35
##                    3rd Qu.:1   3rd Qu.:29.00   3rd Qu.:12.00
##                    Max.   :1   Max.   :48.00   Max.   :16.00
##      black             hisp              marr           nodegree
##  Min.   :0.0000   Min.   :0.00000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:1.0000   1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :1.0000   Median :0.00000   Median :0.0000   Median :1.0000
##  Mean   :0.8432   Mean   :0.05946   Mean   :0.1892   Mean   :0.7081
##  3rd Qu.:1.0000   3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.00000   Max.   :1.0000   Max.   :1.0000
##       re74            re75            re78
##  Min.   :    0   Min.   :    0   Min.   :    0.0
##  1st Qu.:    0   1st Qu.:    0   1st Qu.:  485.2
##  Median :    0   Median :    0   Median : 4232.3
##  Mean   : 2096   Mean   : 1532   Mean   : 6349.1
##  3rd Qu.: 1291   3rd Qu.: 1817   3rd Qu.: 9643.0
##  Max.   :35040   Max.   :25142   Max.   :60307.9
```

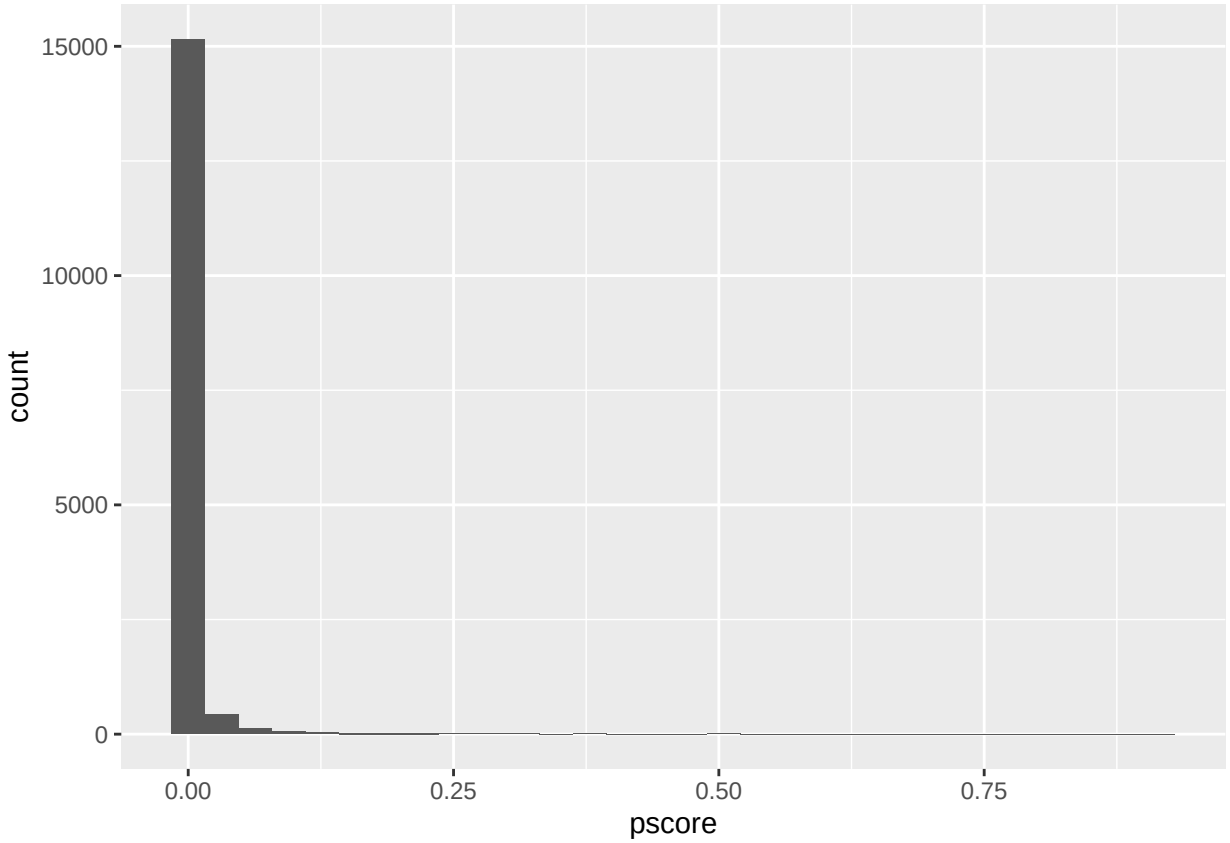
```
mean1 <- nsw_dw %>%
  filter(treat == 1) %>%
  pull(re78) %>%
  mean()

nsw_dw$y1 <- mean1

nsw_dw %>%
  filter(treat == 0) %>%
  summary(re78)
```

```
##    data_id              treat        age              educ
##  Length:260         Min.   :0   Min.   :17.00   Min.   : 3.00
##  Class :character   1st Qu.:0   1st Qu.:19.00   1st Qu.: 9.00
##  Mode  :character   Median :0   Median :24.00   Median :10.00
##                     Mean   :0   Mean   :25.05   Mean   :10.09
##                     3rd Qu.:0   3rd Qu.:28.00   3rd Qu.:11.00
##                     Max.   :0   Max.   :55.00   Max.   :14.00
##      black             hisp              marr           nodegree
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:1.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:1.0000
##  Median :1.0000   Median :0.0000   Median :0.0000   Median :1.0000
##  Mean   :0.8269   Mean   :0.1077   Mean   :0.1538   Mean   :0.8346
##  3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##       re74             re75             re78            y1
##  Min.   :    0.0   Min.   :    0.0   Min.   :    0   Min.   :6349
##  1st Qu.:    0.0   1st Qu.:    0.0   1st Qu.:    0   1st Qu.:6349
##  Median :    0.0   Median :    0.0   Median : 3139   Median :6349
##  Mean   : 2107.0   Mean   : 1266.9   Mean   : 4555   Mean   :6349
##  3rd Qu.:  139.4   3rd Qu.:  650.1   3rd Qu.: 7288   3rd Qu.:6349
##  Max.   :39570.7   Max.   :23032.0   Max.   :39484   Max.   :6349
```

```
mean0 <- nsw_dw %>%
  filter(treat == 0) %>%
```

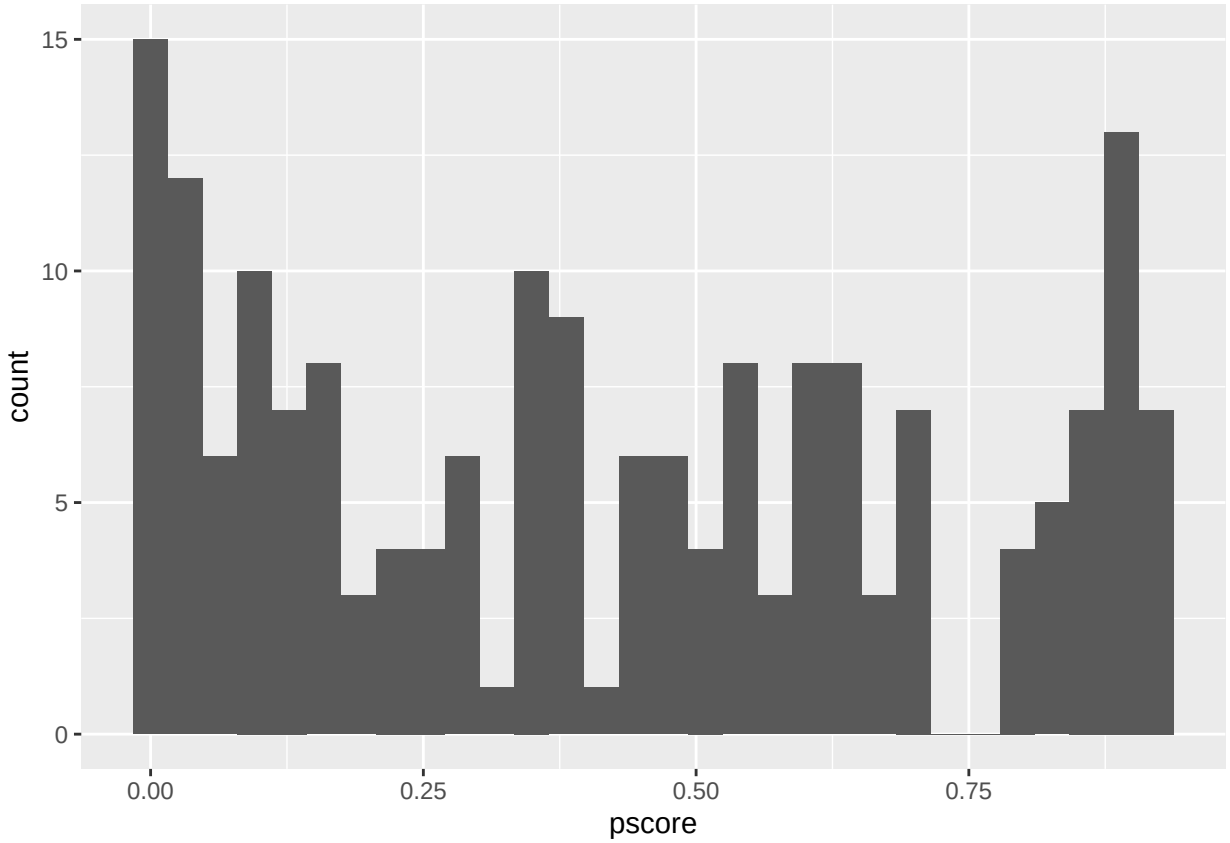
```
  pull(re78) %>%
  mean()

nsw_dw$y0 <- mean0

ate <- unique(nsw_dw$y1 - nsw_dw$y0) # Before and after first difference for each unit?

nsw_dw <- nsw_dw %>%
  filter(treat == 1) %>%
  select(-y1, -y0)

nsw_dw_cpscontrol <- read_data("cps_mixtape.dta") %>%
  bind_rows(nsw_dw) %>%
  mutate(agesq = age^2,
         agecube = age^3,
         educsq = educ*educ,
         u74 = case_when(re74 == 0 ~ 1, TRUE ~ 0),
         u75 = case_when(re75 == 0 ~ 1, TRUE ~ 0),
         interaction1 = educ*re74,
         re74sq = re74^2,
         re75sq = re75^2,
         interaction2 = u74*hisp)


logit_nsw_cube <- glm(treat ~ age + agesq + agecube + educ + educsq + marr + nodegree + black + hisp + i

nsw_dw_cpscontrol <- nsw_dw_cpscontrol %>%
  mutate(pscore = logit_nsw_cube$fitted.values)

pscore_control <- nsw_dw_cpscontrol %>%
  filter(treat == 0) %>%
  pull(pscore) %>%
  mean()

pscore_treated <- nsw_dw_cpscontrol %>%
  filter(treat == 1) %>%
  pull(pscore) %>%
  mean()

nsw_dw_cpscontrol %>%
  filter(treat == 0) %>%
  ggplot() +
  geom_histogram(aes(x = pscore))
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
nsw_dw_cpscontrol %>%
  filter(treat == 1) %>%
  ggplot() +
  geom_histogram(aes(x = pscore))
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
### Max and Min values of the scores for treated and control
nsw_dw_cpscontroltreated <- nsw_dw_cpscontrol %>%
  filter(treat == 1)

max(nsw_dw_cpscontroltreated$pscore)
```

```
## [1] 0.9225552
```

```
min(nsw_dw_cpscontroltreated$pscore)
```

```
## [1] 0.0009875993
```

```
nsw_dw_cpscontroluntreated<- nsw_dw_cpscontrol %>%
  filter(treat == 0)

max(nsw_dw_cpscontroluntreated$pscore)
```

```
## [1] 0.9142296
```

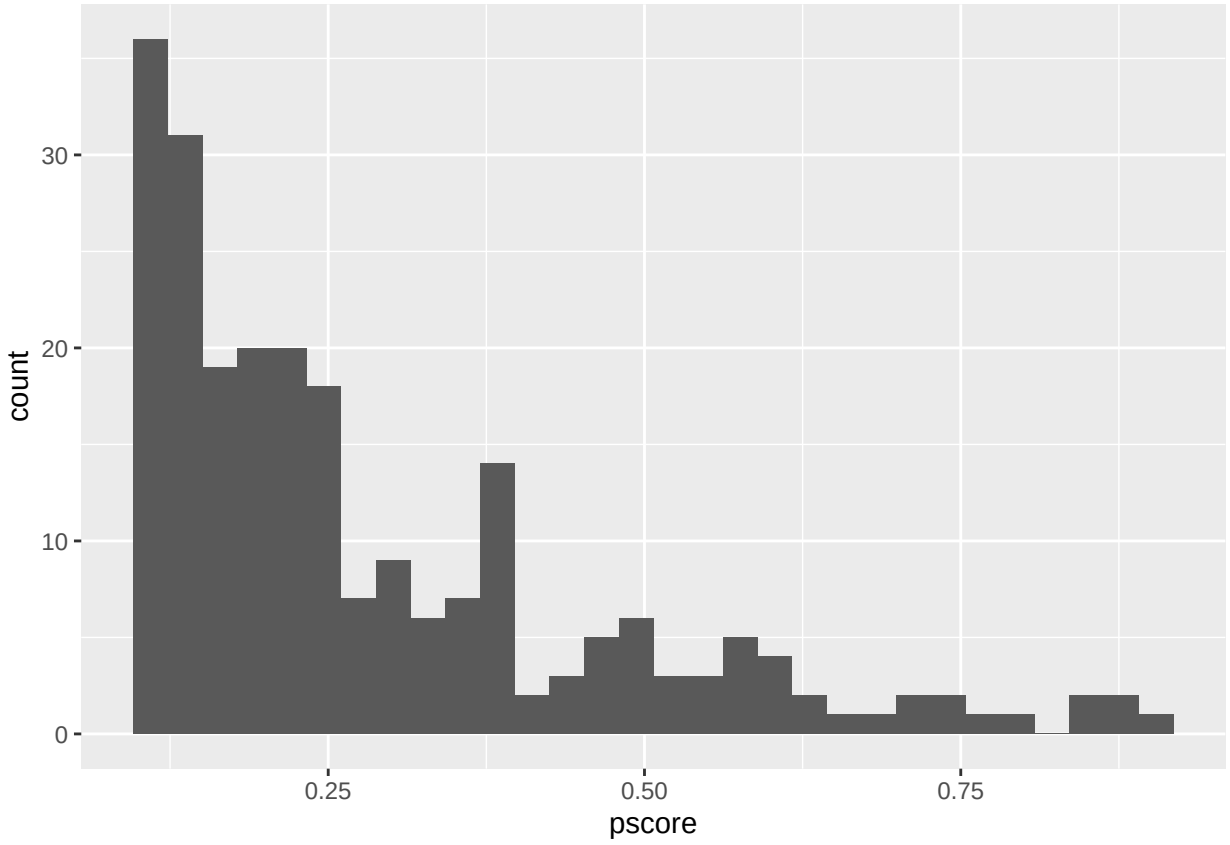
```
min(nsw_dw_cpscontroluntreated$pscore)
```

```
## [1] 1.151399e-08
```

Droping propensity score

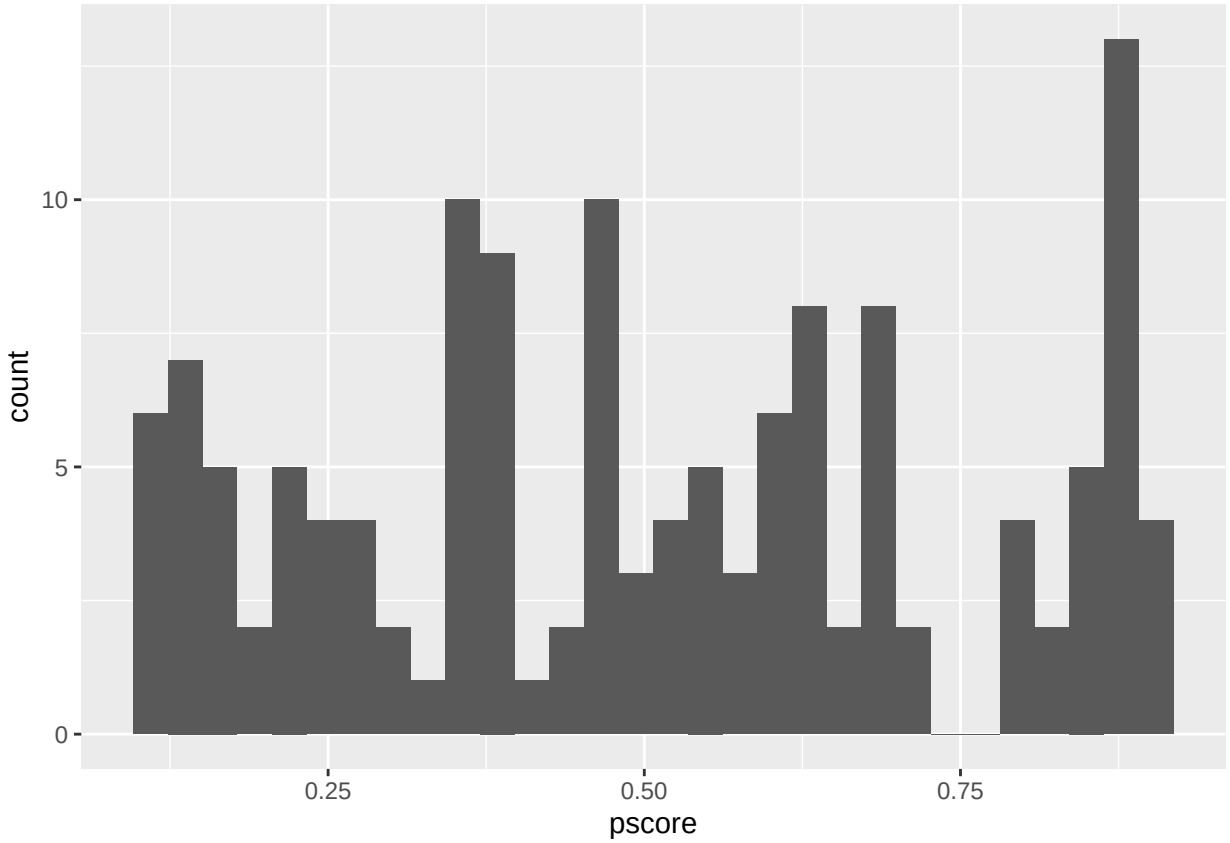
```
nsw_dw_cpscontrolcut <- nsw_dw_cpscontrol %>%
  filter(!(pscore >= 0.9)) %>%
  filter(!(pscore <= 0.1))

nsw_dw_cpscontrolcut %>%
  filter(treat == 0) %>%
  ggplot() +
  geom_histogram(aes(x = pscore))
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
nsw_dw_cpscontrolcut %>%
  filter(treat == 1) %>%
  ggplot() +
  geom_histogram(aes(x = pscore))
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
nsw_dw_cpscontroltreatedcut <- nsw_dw_cpscontrolcut %>%
  filter(treat == 1)

max(nsw_dw_cpscontroltreatedcut$pscore)
```

```
## [1] 0.8966157
```

```
min(nsw_dw_cpscontroltreatedcut$pscore)
```

```
## [1] 0.1016384
```

```
nsw_dw_cpscontroluntreatedcut <- nsw_dw_cpscontrolcut %>%
  filter(treat == 0)

max(nsw_dw_cpscontroluntreatedcut$pscore)
```

```
## [1] 0.8948299
```

```
min(nsw_dw_cpscontroluntreatedcut$pscore)
```

```
## [1] 0.1000871
```

```r
N <- nrow(nsw_dw_cpscontrol)

#- Manual with non-normalized weights using all data
nsw_dw_cpscontrol <- nsw_dw_cpscontrol %>%
  mutate(d1 = treat/pscore,
         d0 = (1-treat)/(1-pscore))
s1 <- sum(nsw_dw_cpscontrol$d1)
s0 <- sum(nsw_dw_cpscontrol$d0)


nsw_dw_cpscontrol <- nsw_dw_cpscontrol %>%
  mutate(y1 = treat * re78/pscore,
         y0 = (1-treat) * re78/(1-pscore),
         ht = y1 - y0)

#- Manual with normalized weights
nsw_dw_cpscontrol <- nsw_dw_cpscontrol %>%
  mutate(y1 = (treat*re78/pscore)/(s1/N),
         y0 = ((1-treat)*re78/(1-pscore))/(s0/N),
         norm = y1 - y0)

nsw_dw_cpscontrol %>%
  pull(ht) %>%
  mean()
```

```
## [1] -11682.31
```

```r
nsw_dw_cpscontrol %>%
  pull(norm) %>%
  mean()
```

```
## [1] -6784.387
```

```r
nsw_dw_cpscontrol %>% pull(norm) %>% mean() - nsw_dw_cpscontrol %>% pull(ht) %>% mean()
```

```
## [1] 4897.924
```

```r
# Trimming data
nsw_dw_cpscontrol <- nsw_dw_cpscontrol %>%
  select(-d1, -d0, -y1, -y0, -ht, -norm) %>%
  filter(!(pscore >= 0.9)) %>%
  filter(!(pscore <= 0.1))

N <- nrow(nsw_dw_cpscontrol)

#- Manual with non-normalized weights using trimmed data
nsw_dw_cpscontrol <- nsw_dw_cpscontrol %>%
  mutate(d1 = treat/pscore,
         d0 = (1-treat)/(1-pscore))

s1 <- sum(nsw_dw_cpscontrol$d1)
```

```
s0 <- sum(nsw_dw_cpscontrol$d0)

nsw_dw_cpscontrol <- nsw_dw_cpscontrol %>%
  mutate(y1 = treat * re78/pscore,
         y0 = (1-treat) * re78/(1-pscore),
         ht = y1 - y0)

#- Manual with normalized weights with trimmed data
nsw_dw_cpscontrol <- nsw_dw_cpscontrol %>%
  mutate(y1 = (treat*re78/pscore)/(s1/N),
         y0 = ((1-treat)*re78/(1-pscore))/(s0/N),
         norm = y1 - y0)

nsw_dw_cpscontrol %>%
  pull(ht) %>%
  mean()
```

```
## [1] 1551.058
```

```
nsw_dw_cpscontrol %>%
  pull(norm) %>%
  mean()
```

```
## [1] 1350.894
```

```
nsw_dw_cpscontrol %>% pull(norm) %>% mean() - nsw_dw_cpscontrol %>% pull(ht) %>% mean()
```

```
## [1] -200.1645
```

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.