Complete_RDD2

Joey Herrera

4/16/2021

Causal Inference Replication 2

Load data from the RDD2 repository and add addtional variables

```
# Create a path to read data from the source
read_data <- function(df)
{
  full_path <- paste("https://raw.github.com/Joey-Herrera/Causal_inference/main/RDD2/Data/",
                     df, sep = "")
  df <- read_dta(full_path)
  return(df)
}

nsw_dw <- read_dta("https://raw.github.com/Joey-Herrera/Causal_inference/main/RDD2/Data/nsw_mixtape.dta

nsw_dw <- nsw_dw %>%
  filter(treat == 1)

# Append variables as necessary
nsw_lm <- read_dta("https://raw.github.com/Joey-Herrera/Causal_inference/main/RDD2/Data/cps_mixtape.dta
  bind_rows(nsw_dw) %>%
  mutate(agesq = age^2,
         agecube = age^3,
         educsq = educ*educ,
         educcube = educ^3,
         u74 = case_when(re74 == 0 ~ 1, TRUE ~ 0),
         u75 = case_when(re75 == 0 ~ 1, TRUE ~ 0),
         re74sq = re74^2,
         re74cube = re74^3,
         re75sq = re75^2,
         re75cube = re75^3,
         re78sq = re78^2,
         re75cube = re78^3,
          interaction1 = educ*re74,
         interaction2 = u74*hisp)

nsw_logit <- read_dta("https://raw.github.com/Joey-Herrera/Causal_inference/main/RDD2/Data/cps_mixtape.
  bind_rows(nsw_dw) %>%
  mutate(agesq = age^2,
         agecube = age^3,
         educsq = educ*educ,
         educcube = educ^3,
```

```
         u74 = case_when(re74 == 0 ~ 1, TRUE ~ 0),
         u75 = case_when(re75 == 0 ~ 1, TRUE ~ 0),
         re74sq = re74^2,
         re74cube = re74^3,
         re75sq = re75^2,
         re75cube = re75^3,
         re78sq = re78^2,
         re75cube = re78^3,
         interaction1 = educ*re74,
         interaction2 = u74*hisp)
```

Question 1

```
# Question 1A
# fitting model using a logit regression
logit_nsw <- glm(treat ~ age + agesq + agecube + educ + educsq +
                   marr + nodegree + black + hisp + re74 + re75 + u74 +
                   u75 + interaction1, family = binomial(link = "logit"),
                 data = nsw_logit)
logit_nsw
```

```
##
## Call:  glm(formula = treat ~ age + agesq + agecube + educ + educsq +
##     marr + nodegree + black + hisp + re74 + re75 + u74 + u75 +
##     interaction1, family = binomial(link = "logit"), data = nsw_logit)
##
## Coefficients:
##  (Intercept)           age         agesq       agecube          educ
##   -3.522e+01     2.425e+00    -6.724e-02     5.685e-04     9.248e-01
##       educsq          marr      nodegree         black          hisp
##   -5.720e-02    -1.556e+00     9.271e-01     3.851e+00     1.674e+00
##         re74          re75           u74           u75  interaction1
##   -2.203e-04    -1.969e-04     1.750e+00     9.440e-03     2.222e-05
##
## Degrees of Freedom: 16176 Total (i.e. Null);  16162 Residual
## Null Deviance:       2022
## Residual Deviance: 808.3     AIC: 838.3
```

```
# estimating model using linear probability model
OLS_nsw <- lm(treat ~ age + agesq + agecube + educ + educsq +
                   marr + nodegree + black + hisp + re74 + re75 + u74 +
                   u75 + interaction1,data = nsw_lm)

OLS_nsw
```

```
##
## Call:
## lm(formula = treat ~ age + agesq + agecube + educ + educsq +
##     marr + nodegree + black + hisp + re74 + re75 + u74 + u75 +
##     interaction1, data = nsw_lm)
##
## Coefficients:
##  (Intercept)           age         agesq       agecube          educ
##   -3.714e-01     3.241e-02    -8.975e-04     7.795e-06     6.558e-03
```

```
##       educsq          marr      nodegree         black          hisp
##   -4.434e-04    -2.352e-02     2.151e-02     1.030e-01     6.795e-03
##         re74          re75          u74           u75  interaction1
##   -3.064e-06    -1.458e-07     4.535e-02     2.094e-02     2.853e-07
```

```
# Question 1B
# Fit a propensity score using a quadratic for every variable and another with a cubic for every variab
####################
##### Logit
### quadratic max
# take out age cubed
logit_model_quad <- glm(treat ~ age + agesq  + educ + educsq +
                  marr + nodegree + black + hisp + re74 + re74sq + re75 + re75sq +
                    u74 + u75, family = binomial(link = "logit"),
                data = nsw_logit)

# Fit propoensity score to quad model
nsw_logit_quad <- nsw_logit %>%
  mutate(pscore = logit_model_quad$fitted.values)

# mean pscore control
pscore_control_logit_quad <- nsw_logit_quad %>%
  filter(treat == 0) %>%
  pull(pscore) %>%
  mean()
# 0.00691

# mean pscore treated
pscore_treated_logit_quad <- nsw_logit_quad %>%
  filter(treat == 1) %>%
  pull(pscore) %>%
  mean()
# 0.4025
####################################
### cube max
logit_model_cube <- glm(treat ~ age + agesq + agecube + educ + educsq + educcube +
                  marr + nodegree + black + hisp + re74 + re74sq + re74cube + re75 + re75sq + re75cube
                  u75, family = binomial(link = "logit"),
                data = nsw_logit)

# Fit propoensity score to cubic model
nsw_logit_cube <- nsw_logit %>%
  mutate(pscore = logit_model_cube$fitted.values)

# mean pscore control
pscore_control_cube <- nsw_logit_cube %>%
  filter(treat == 0) %>%
  pull(pscore) %>%
  mean()
# 0.00662

# mean pscore treated
pscore_treated_cube <- nsw_logit_cube %>%
  filter(treat == 1) %>%
```

```
  pull(pscore) %>%
  mean()
# 0.4279


#########################################################
#Linear probability model
# Quad max
OLS_model_quad <- lm(treat ~ age + agesq + educ + educsq +
                marr + nodegree + black + hisp + re74 + re74sq + re75 + re75sq + u74 +
                u75,data = nsw_lm)

# Fit propoensity score to quad model
nsw_OLS_quad <- nsw_lm %>%
  mutate(pscore = OLS_model_quad$fitted.values)

# mean pscore control
pscore_control_quad <- nsw_OLS_quad %>%
  filter(treat == 0) %>%
  pull(pscore) %>%
  mean()

# 0.00993

# mean pscore treated
pscore_treated_quad <- nsw_OLS_quad %>%
  filter(treat == 1) %>%
  pull(pscore) %>%
  mean()
# 0.14128

#############
# Cube max
OLS_model_cube <- lm(treat ~ age + agesq + agecube + educ + educsq + educcube +
                marr + nodegree + black + hisp + re74 + re74sq + re74cube +
                  re75 + re75sq + re75cube + u74 +
                u75,data = nsw_lm)

# Fit propoensity score to cubic model
nsw_OLS_cube <- nsw_lm %>%
  mutate(pscore = OLS_model_cube$fitted.values)

# mean pscore control
pscore_control_OLScube <- nsw_OLS_cube %>%
  filter(treat == 0) %>%
  pull(pscore) %>%
  mean()

# 0.00985

# mean pscore treated
pscore_treated_OLScube <- nsw_OLS_cube %>%
  filter(treat == 1) %>%
```

```
  pull(pscore) %>%
  mean()
# 0.14828
```

The propensity scores for the four models are as follows:

quadratic linear probability model treatment pscore: 0.14128 control pscore: 0.00993

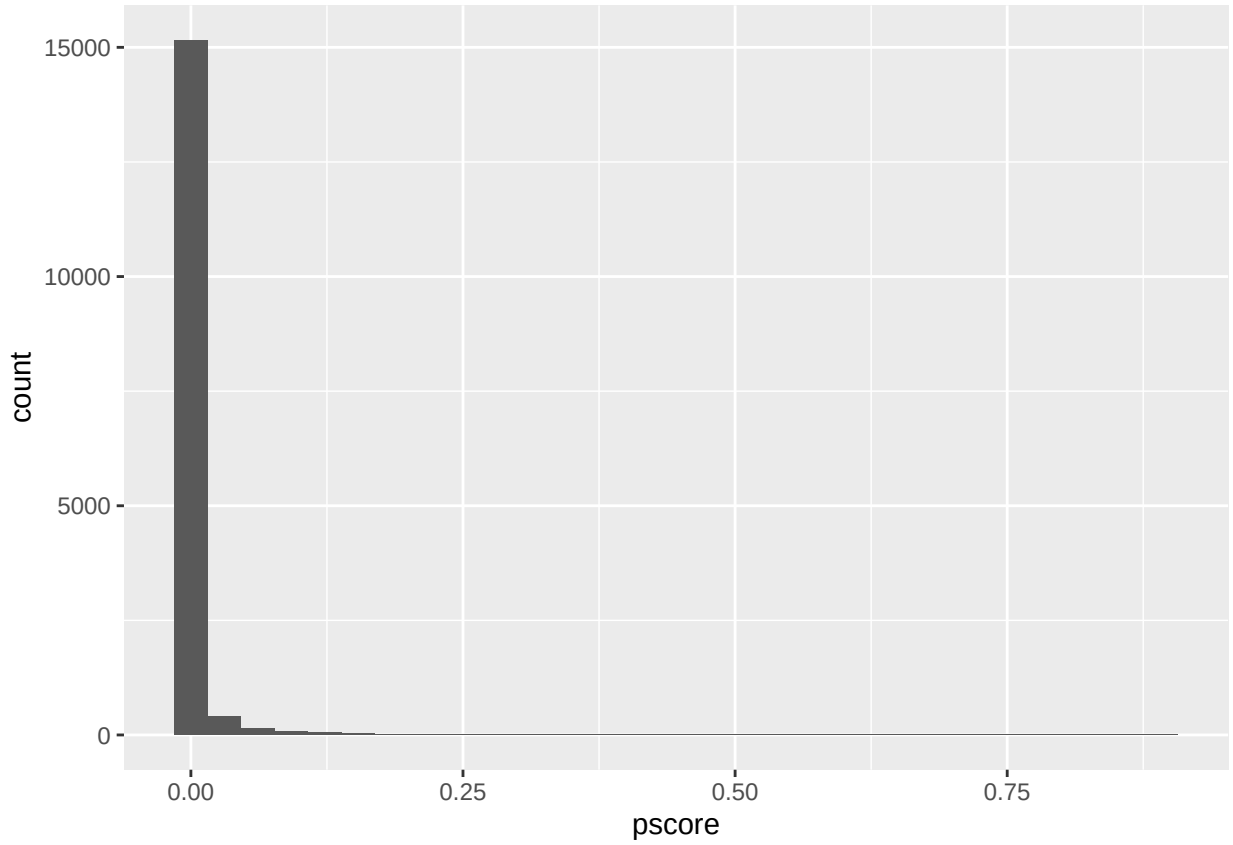
cubic linear probability model treatment pscore: 0.14828 control pscore: 0.00985

quadratic logit model treatment pscore: 0.4025 control pscore: 0.00691

cubic logit model treatment pscore: 0.4279 control pscore: 0.00662

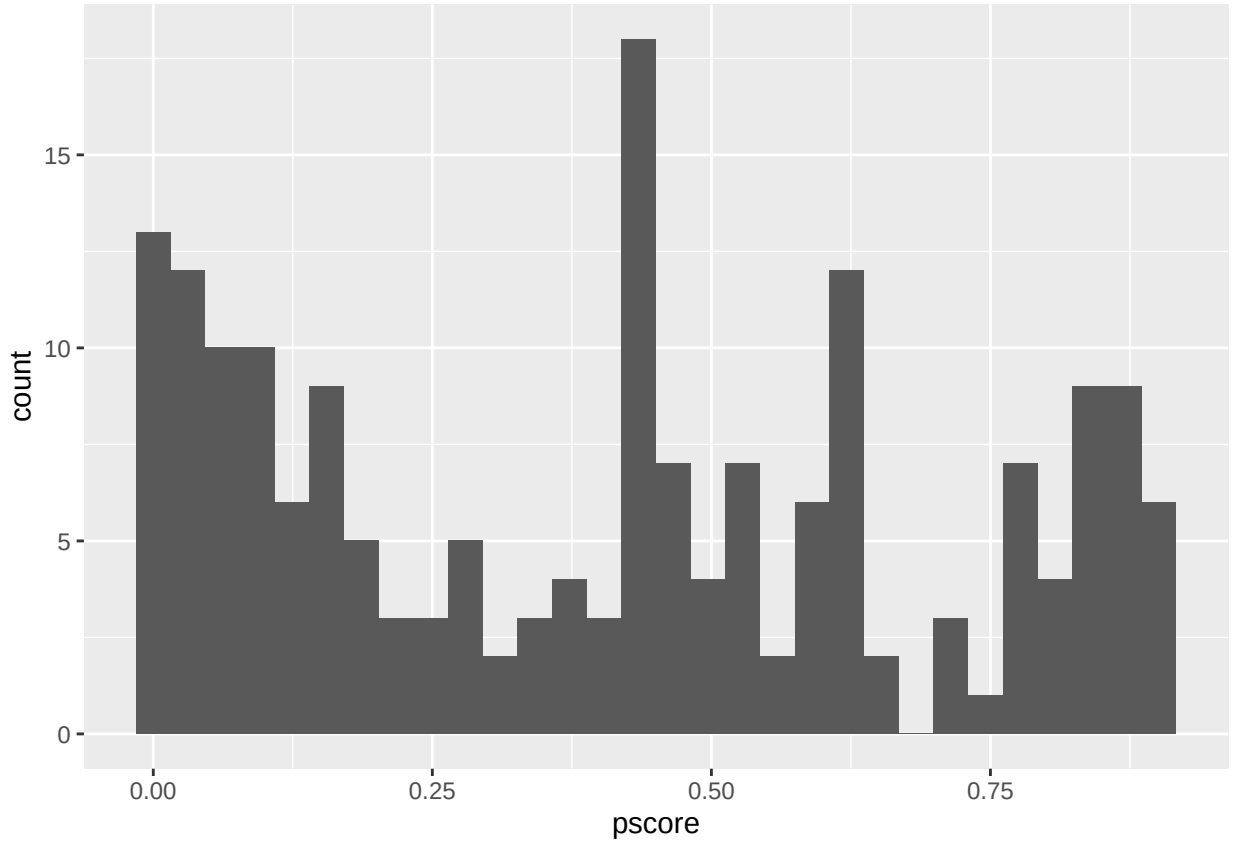
```
##### Question 1C Histograms

# logit quad control
nsw_logit_quad %>%
  filter(treat == 0) %>%
  ggplot() +
  geom_histogram(aes(x = pscore))
```

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



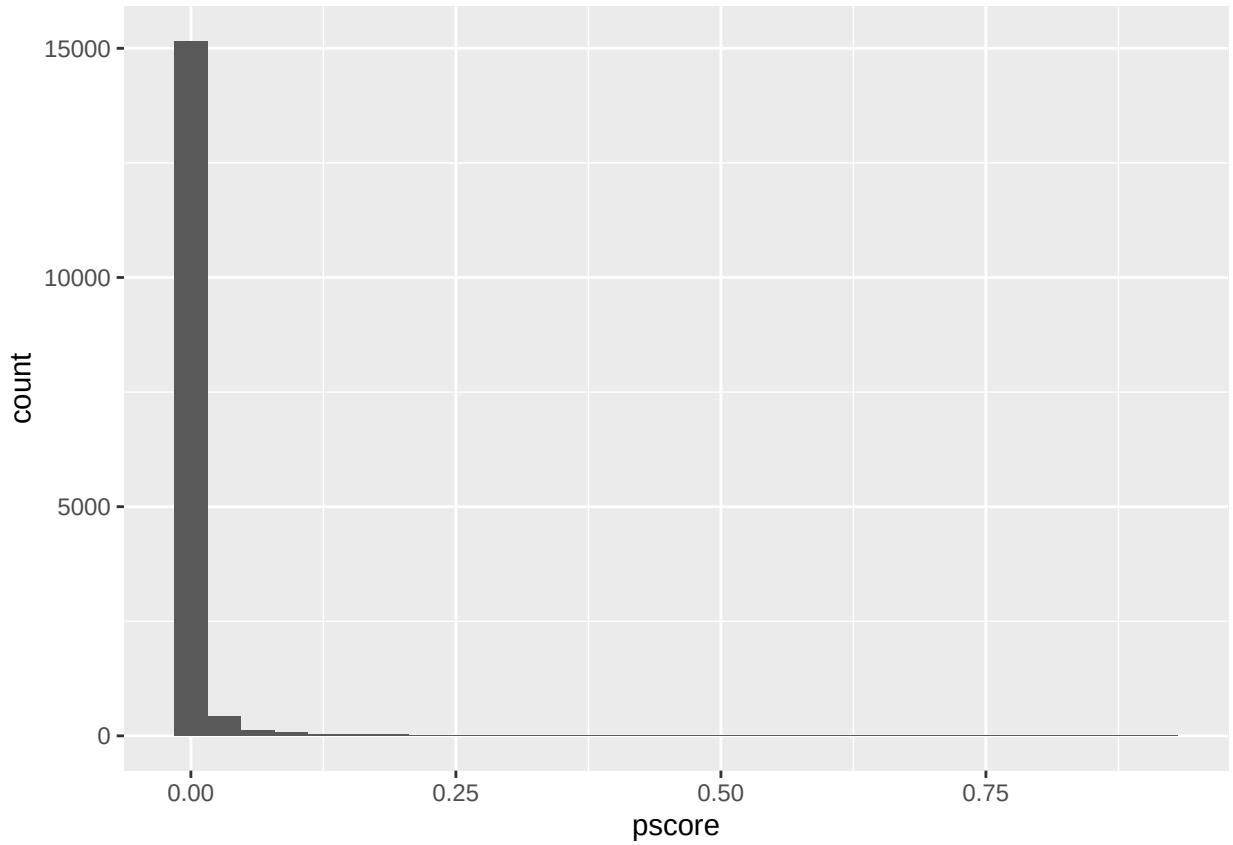

```
# logit quad treatment
nsw_logit_quad %>%
  filter(treat == 1) %>%
  ggplot() +
  geom_histogram(aes(x = pscore))
```

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



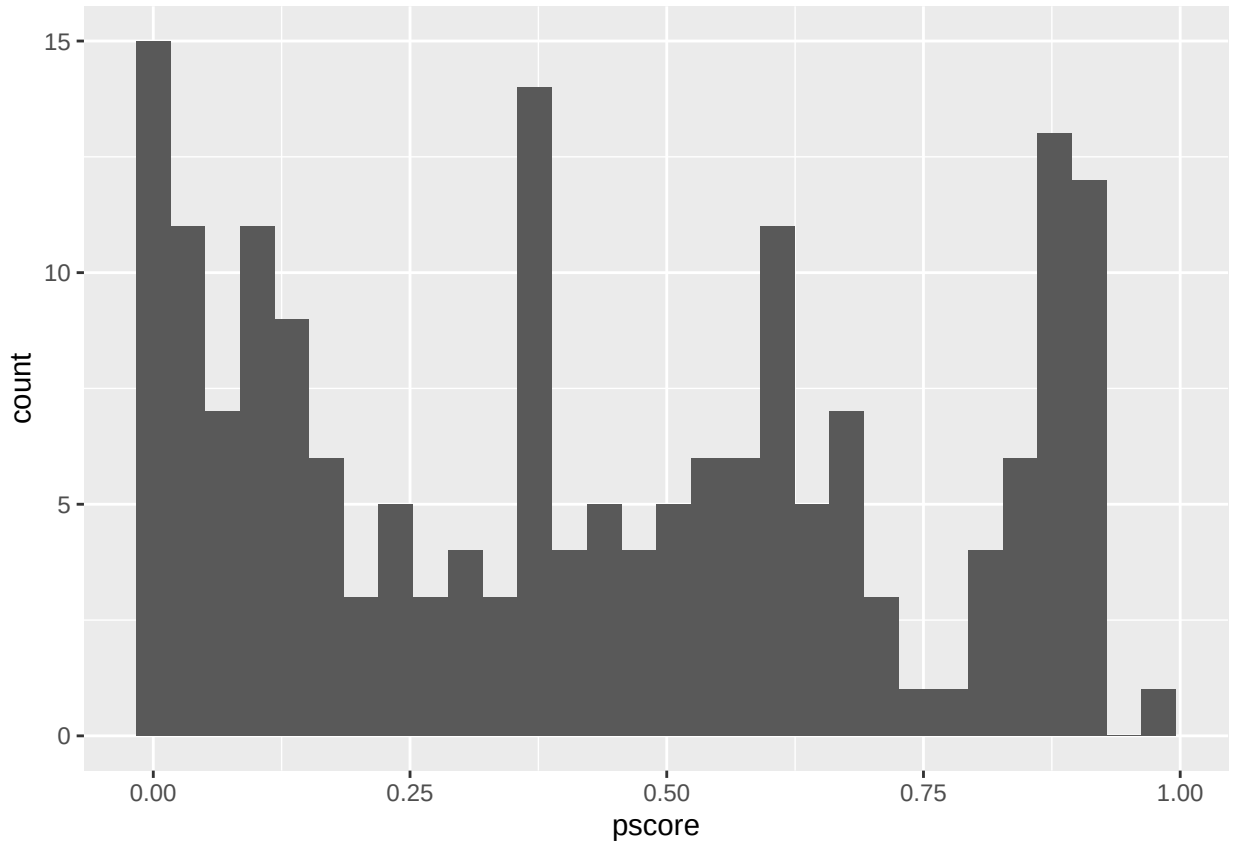

```
# logit cube control
nsw_logit_cube %>%
  filter(treat == 0) %>%
  ggplot() +
  geom_histogram(aes(x = pscore))
```

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



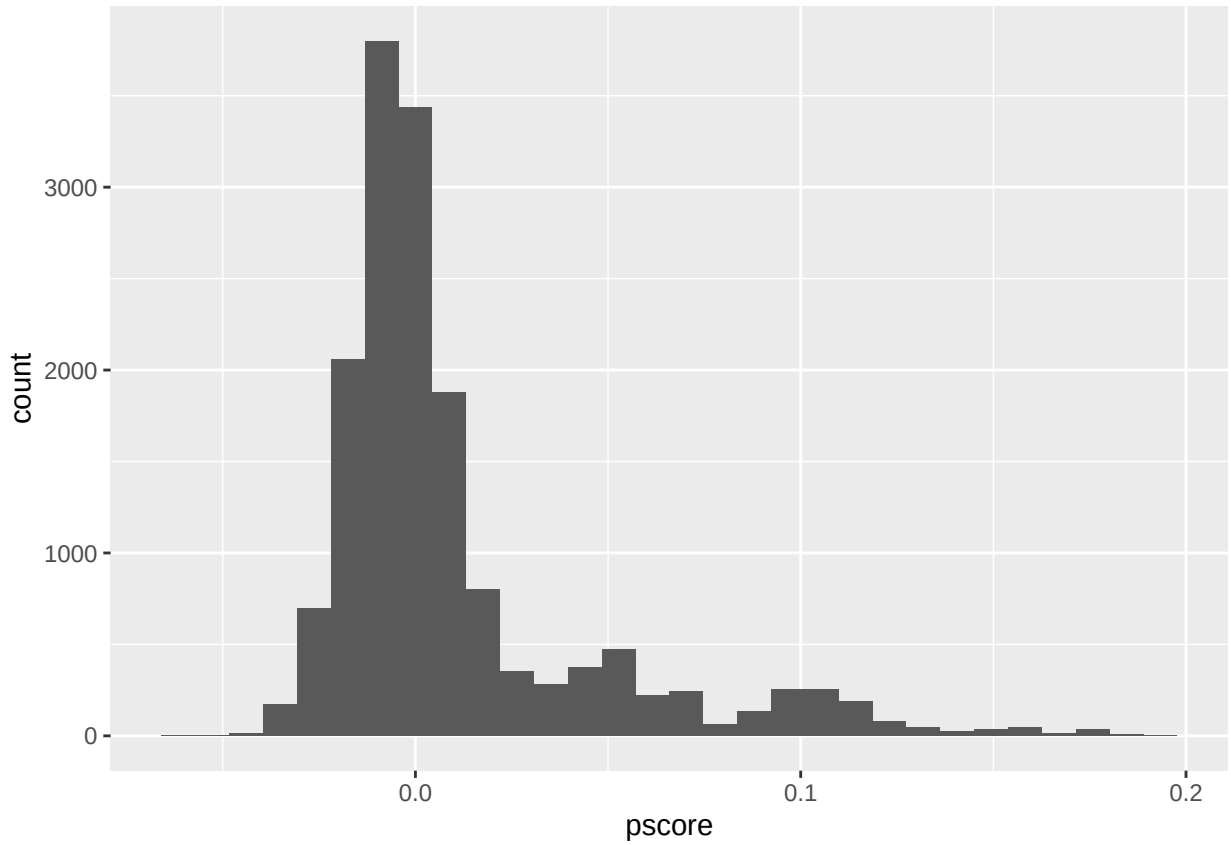

```
# logit cube treatment
nsw_logit_cube %>%
  filter(treat == 1) %>%
  ggplot() +
  geom_histogram(aes(x = pscore))
```

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



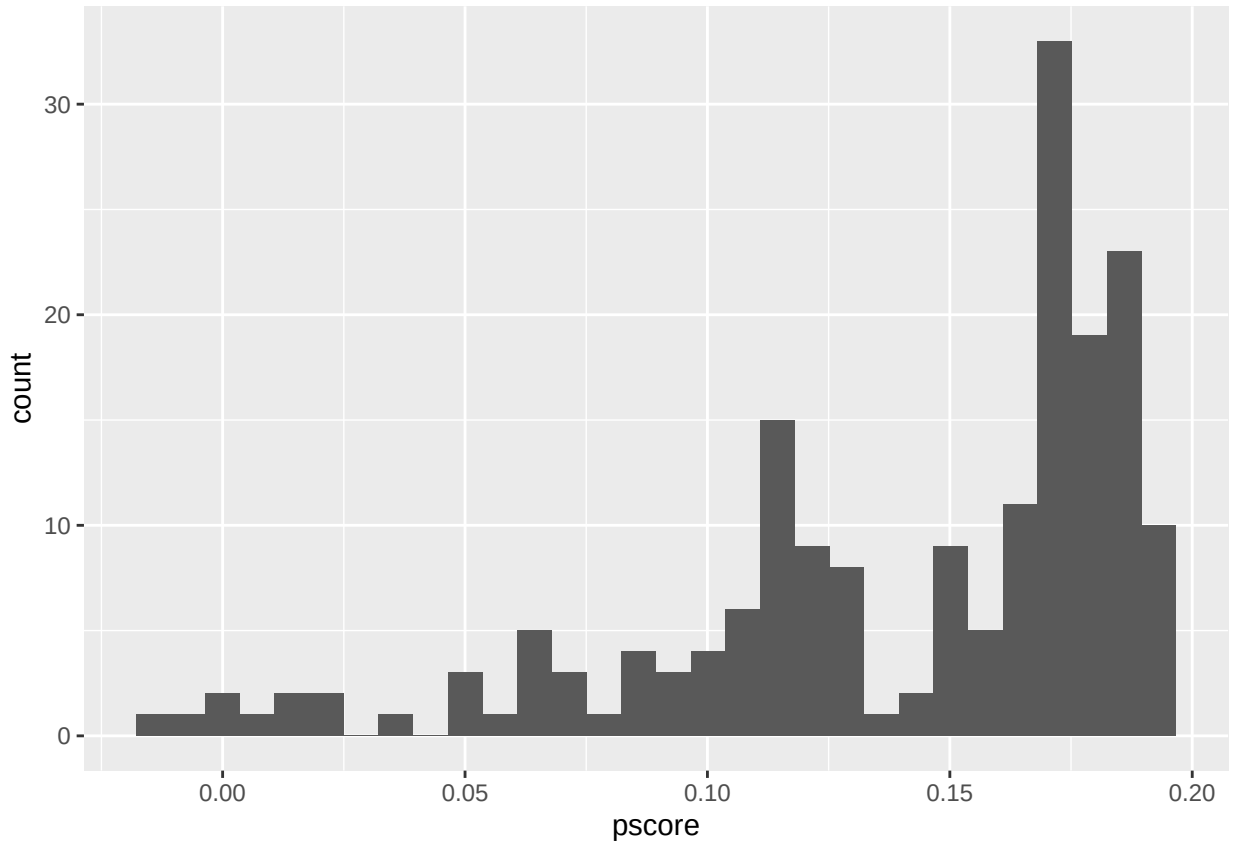

```
# linear probability model (quadratic) control
nsw_OLS_quad %>%
  filter(treat == 0) %>%
  ggplot() +
  geom_histogram(aes(x = pscore))
```

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



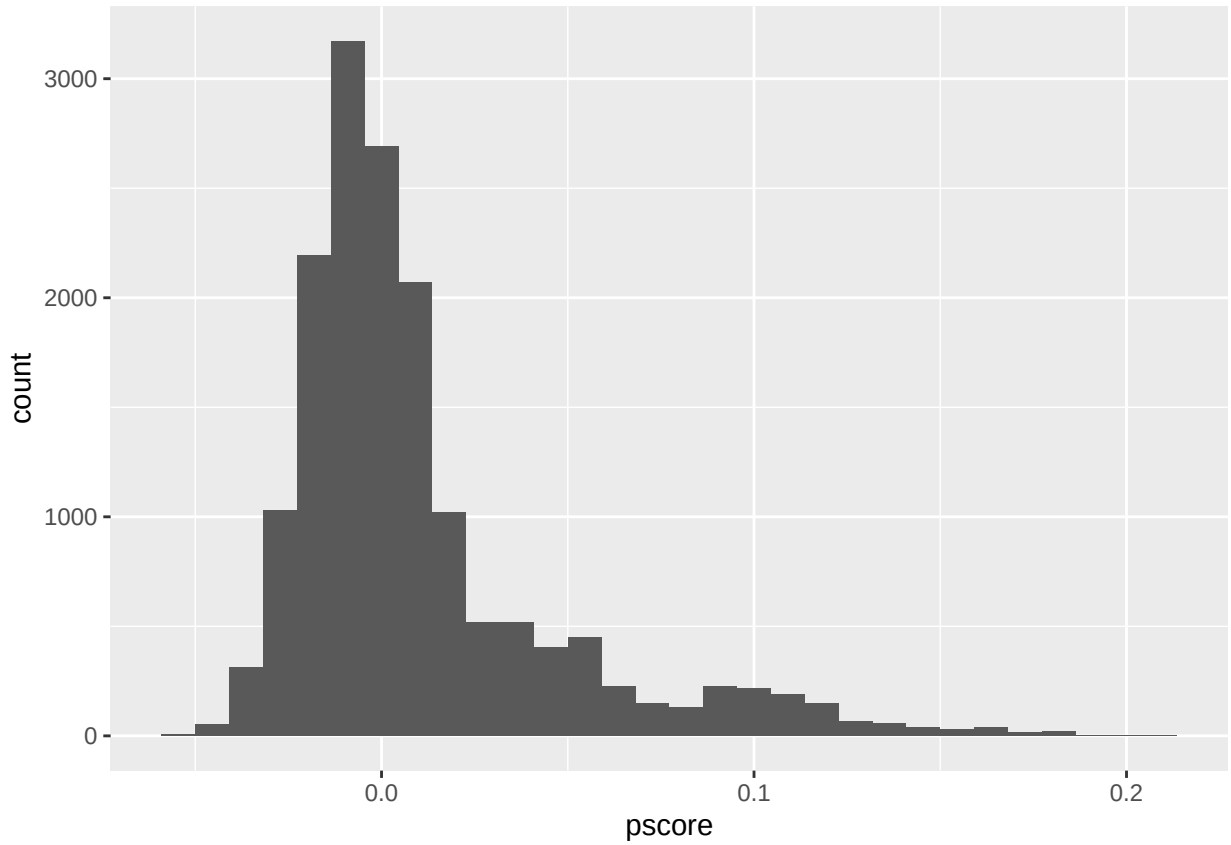

```
# linear probability model (quadratic) treatment
nsw_OLS_quad %>%
  filter(treat == 1) %>%
  ggplot() +
  geom_histogram(aes(x = pscore))
```

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



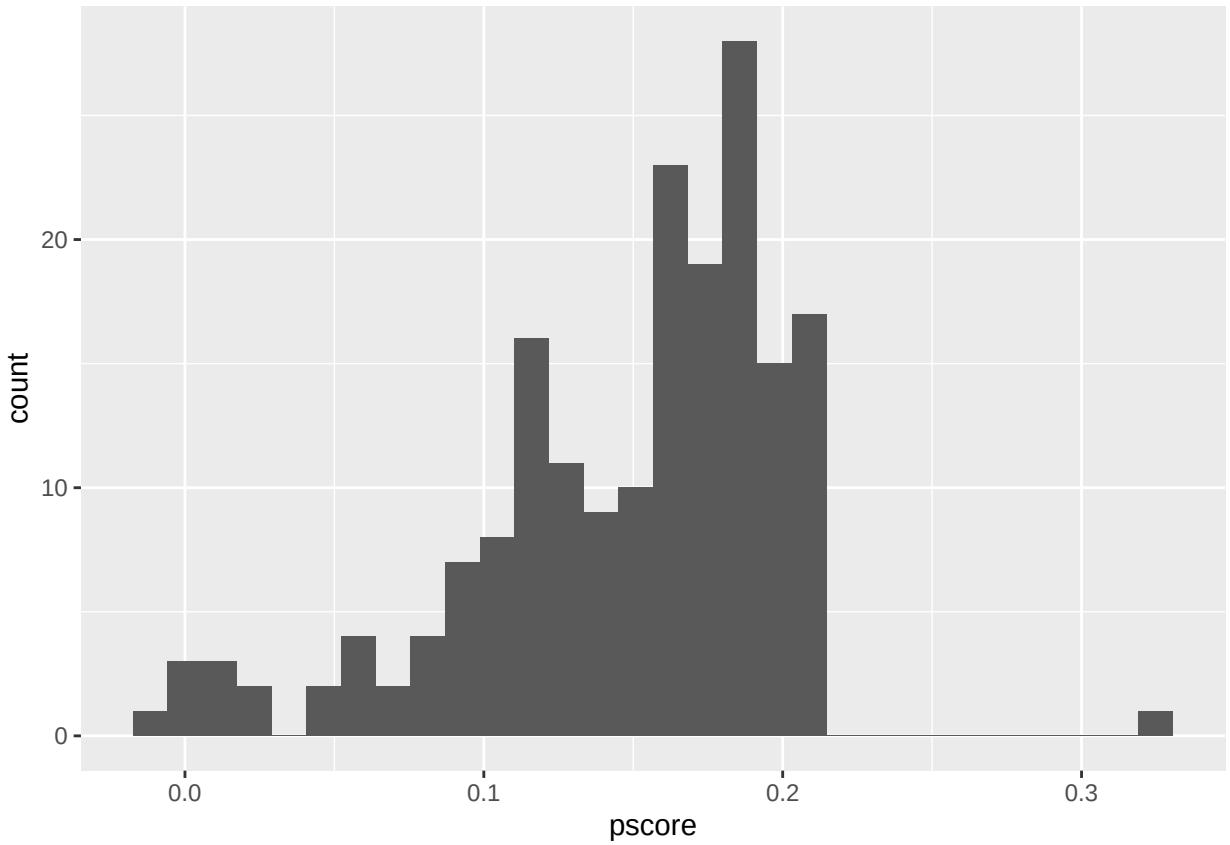

```
# linear probability model (cubic) control
nsw_OLS_cube %>%
  filter(treat == 0) %>%
  ggplot() +
  geom_histogram(aes(x = pscore))
```

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




```
# linear probability model (cubic) treatment
nsw_OLS_cube %>%
  filter(treat == 1) %>%
  ggplot() +
  geom_histogram(aes(x = pscore))
```

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



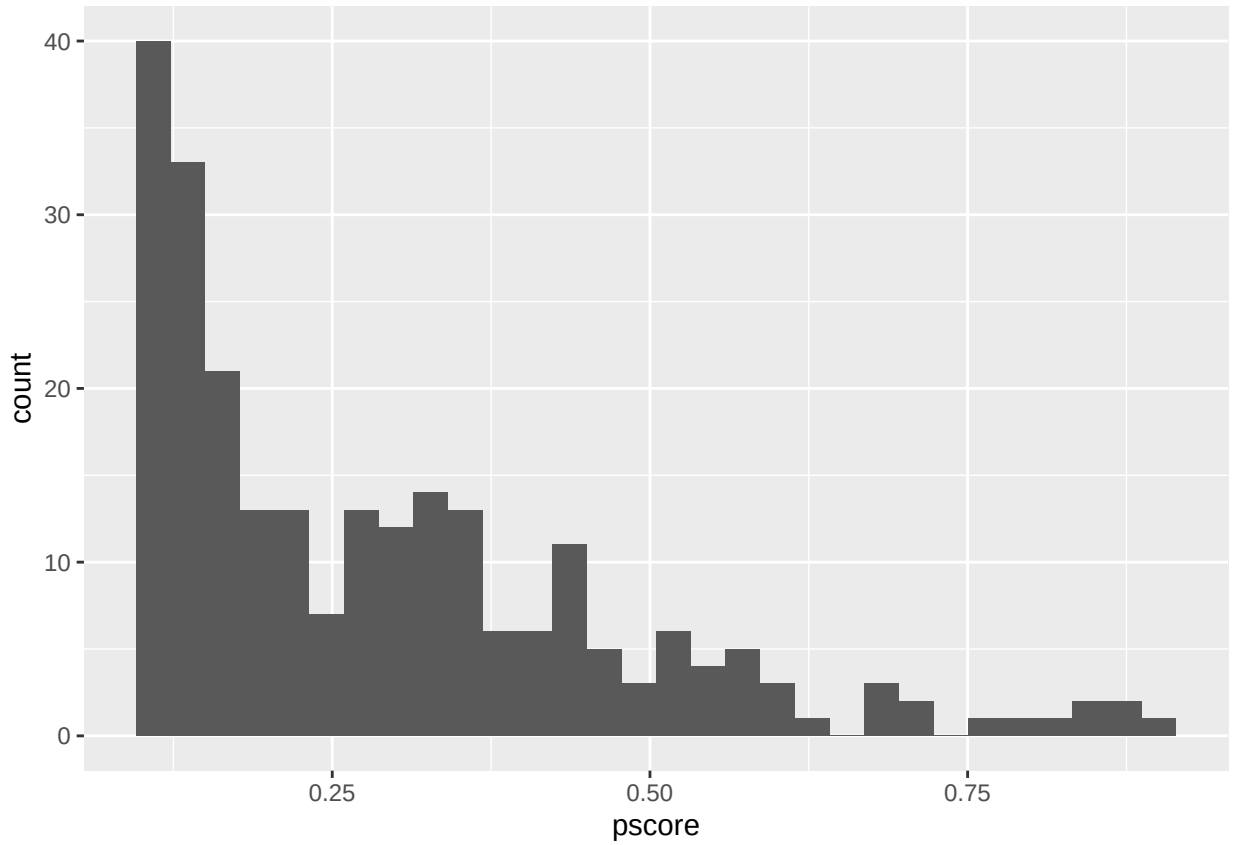
The histograms for the control groups for the logistic models have minimum pscores of 0 and a maximum pscore of about 0.4. The treatment groups for the logistic models have treatment groups with minimum pscores of 0 and a maximum pscore of 1. The linear proability models have control groups with minimum pscores of about -0.05 and a maximum of 0.2. The treatment groups have minimum pscores of -0.02 and a maximum of about 0.2.

```
# Question 1D. trim the propsenity scores and recreate the histograms from 1C

# trimming propensity score for logit and OLS data
# logit model (quadratic)
nsw_logit_quad_trim <- nsw_logit_quad %>%
  filter(!(pscore >= 0.9)) %>%
  filter(!(pscore <= 0.1))

# logit model (cubic)
nsw_logit_cube_trim <- nsw_logit_cube %>%
  filter(!(pscore >= 0.9)) %>%
  filter(!(pscore <= 0.1))

# linear probability model (quadratic)
nsw_OLS_quad_trim <- nsw_OLS_quad %>%
  filter(!(pscore >= 0.9)) %>%
  filter(!(pscore <= 0.1))

# linear probability model (cubic)
nsw_OLS_cube_trim <- nsw_OLS_cube %>%
  filter(!(pscore >= 0.9)) %>%
  filter(!(pscore <= 0.1))
```

```
#####################################################################
# Rerun Question 1B
##### Logit
### quadratic max

# mean pscore control
pscore_control_logit_quad_trim <- nsw_logit_quad_trim %>%
  filter(treat == 0) %>%
  pull(pscore) %>%
  mean()
# 0.2506

# mean pscore treated
pscore_treated_logit_quad_trim <- nsw_logit_quad_trim %>%
  filter(treat == 1) %>%
  pull(pscore) %>%
  mean()
# 0.2768
####################################
### cube max

# mean pscore control
pscore_control_cube_trim <- nsw_logit_cube_trim %>%
  filter(treat == 0) %>%
  pull(pscore) %>%
  mean()
# 0.2516

# mean pscore treated
pscore_treated_cube_trim <- nsw_logit_cube_trim %>%
  filter(treat == 1) %>%
  pull(pscore) %>%
  mean()
# 0.2902


#############################################################
#Linear probability model
# Quad max
# mean pscore control
pscore_control_quad_trim <- nsw_OLS_quad_trim %>%
  filter(treat == 0) %>%
  pull(pscore) %>%
  mean()

# 0.1286

# mean pscore treated
pscore_treated_quad_trim <- nsw_OLS_quad_trim %>%
  filter(treat == 1) %>%
  pull(pscore) %>%
  mean()
# 0.1374
```

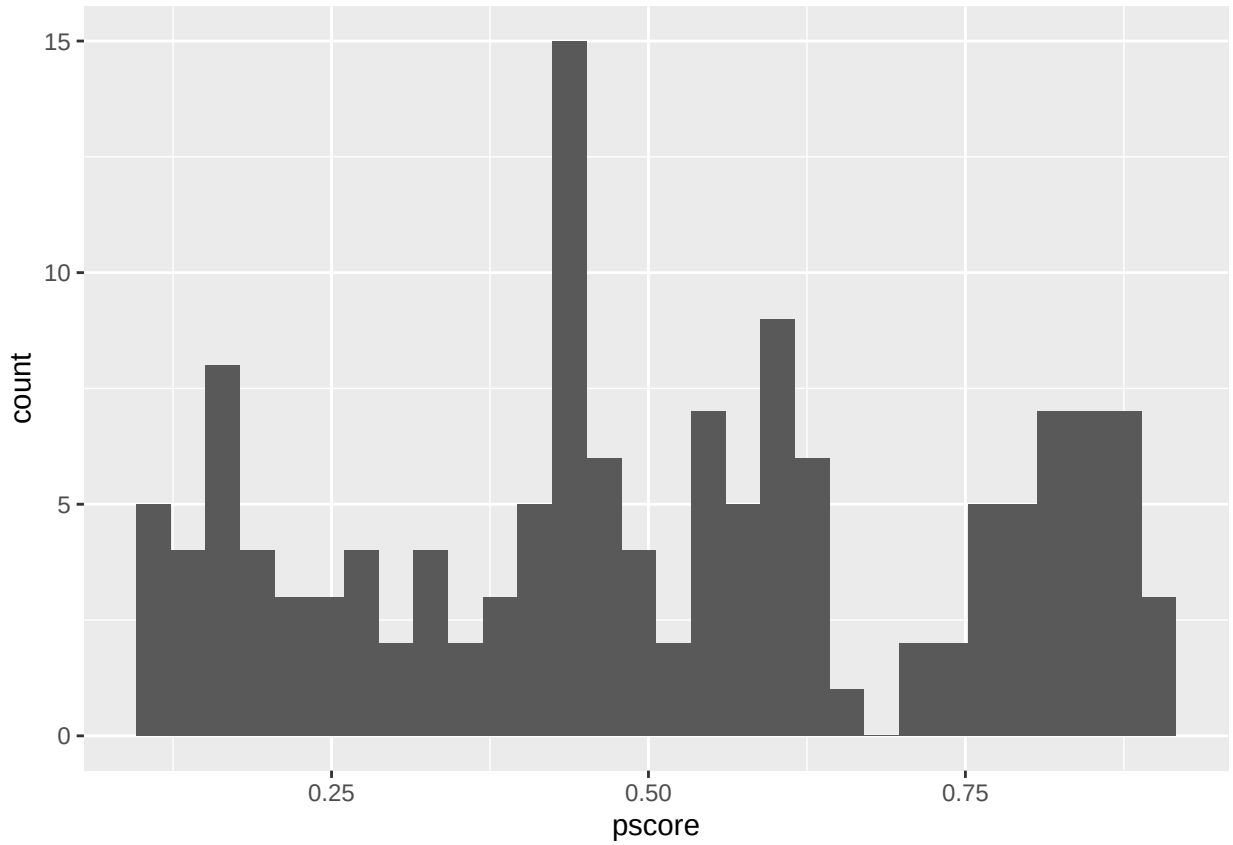
```
############
#linear proability model (cubic)
# mean pscore control
pscore_control_OLS_cube_trim <- nsw_OLS_cube_trim %>%
  filter(treat == 0) %>%
  pull(pscore) %>%
  mean()

# 0.1294

# mean pscore treated
pscore_treated_OLS_cube_trim <- nsw_OLS_cube_trim %>%
  filter(treat == 1) %>%
  pull(pscore) %>%
  mean()
# 0.1404

################################
# rerun the histograms from 1C

# logit quad control
nsw_logit_quad_trim %>%
  filter(treat == 0) %>%
  ggplot() +
  geom_histogram(aes(x = pscore))
```

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



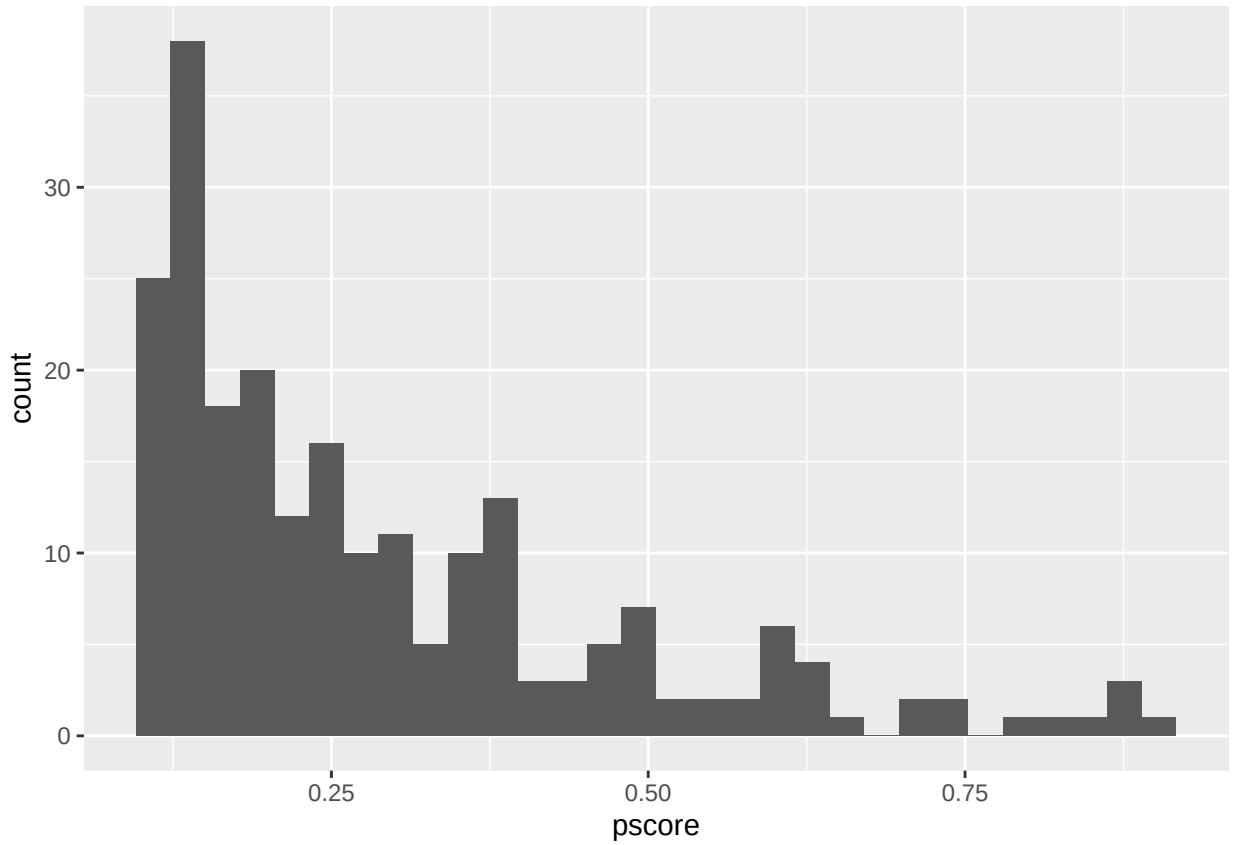

```
# logit quad treatment
nsw_logit_quad_trim %>%
  filter(treat == 1) %>%
  ggplot() +
  geom_histogram(aes(x = pscore))
```

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



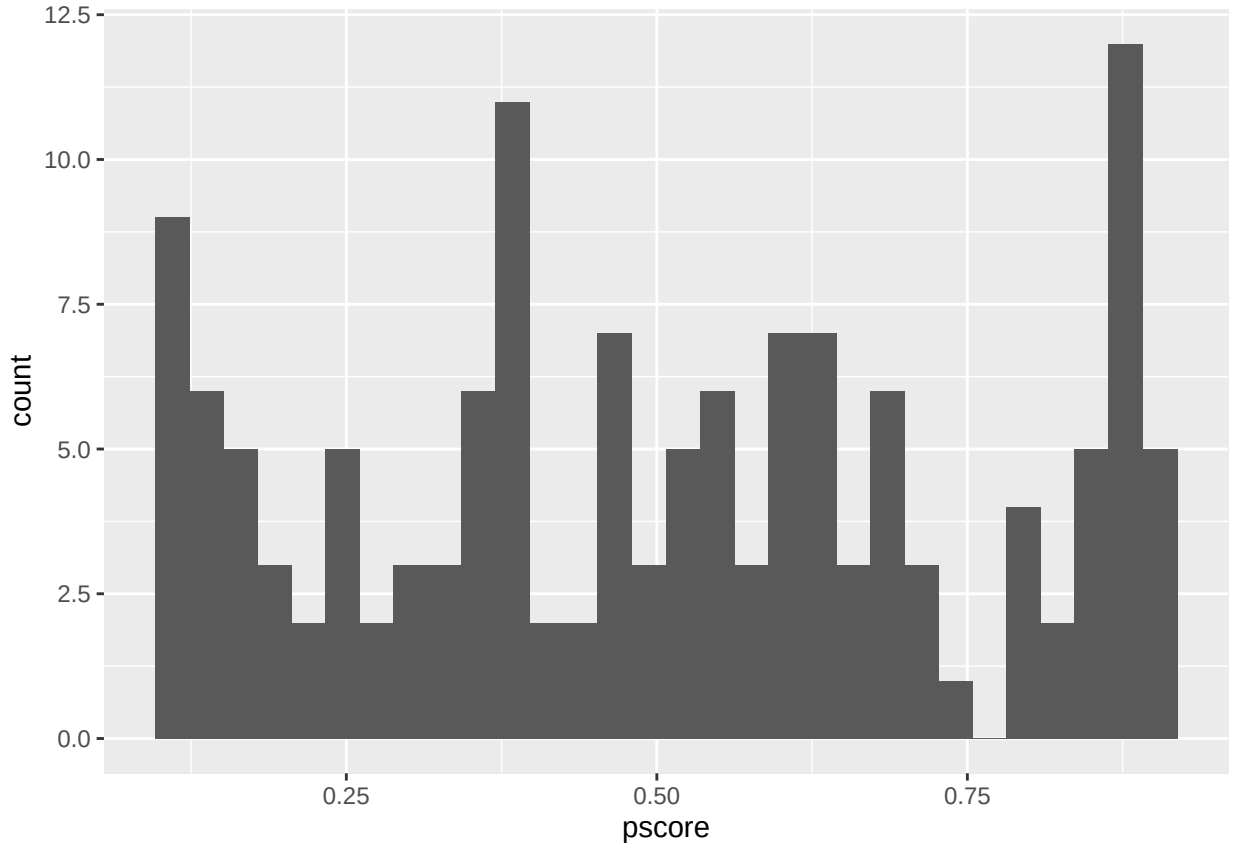

```
# logit cube control
nsw_logit_cube_trim %>%
  filter(treat == 0) %>%
  ggplot() +
  geom_histogram(aes(x = pscore))
```

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



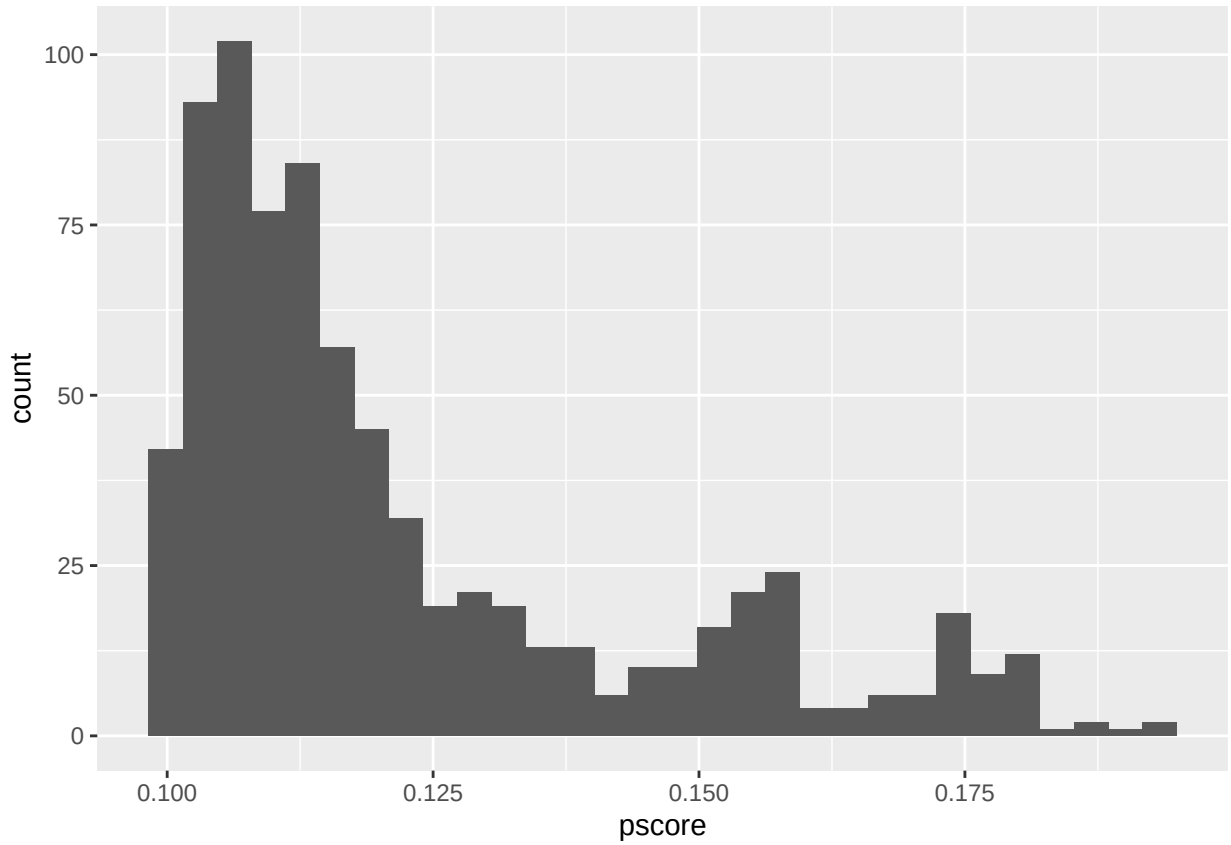

```
# logit cube treatment
nsw_logit_cube_trim %>%
  filter(treat == 1) %>%
  ggplot() +
  geom_histogram(aes(x = pscore))
```

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



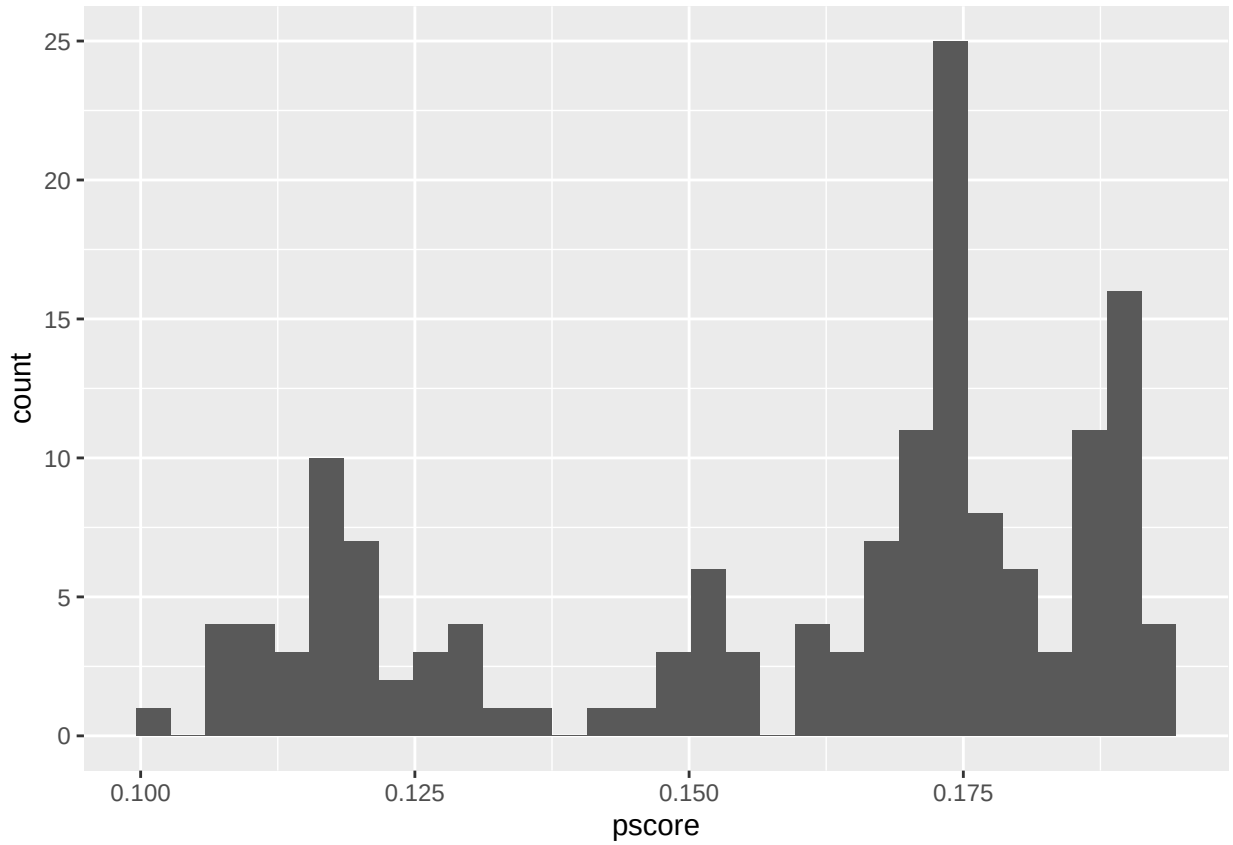

```
# linear probability model (quadratic) control
nsw_OLS_quad_trim %>%
  filter(treat == 0) %>%
  ggplot() +
  geom_histogram(aes(x = pscore))
```

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



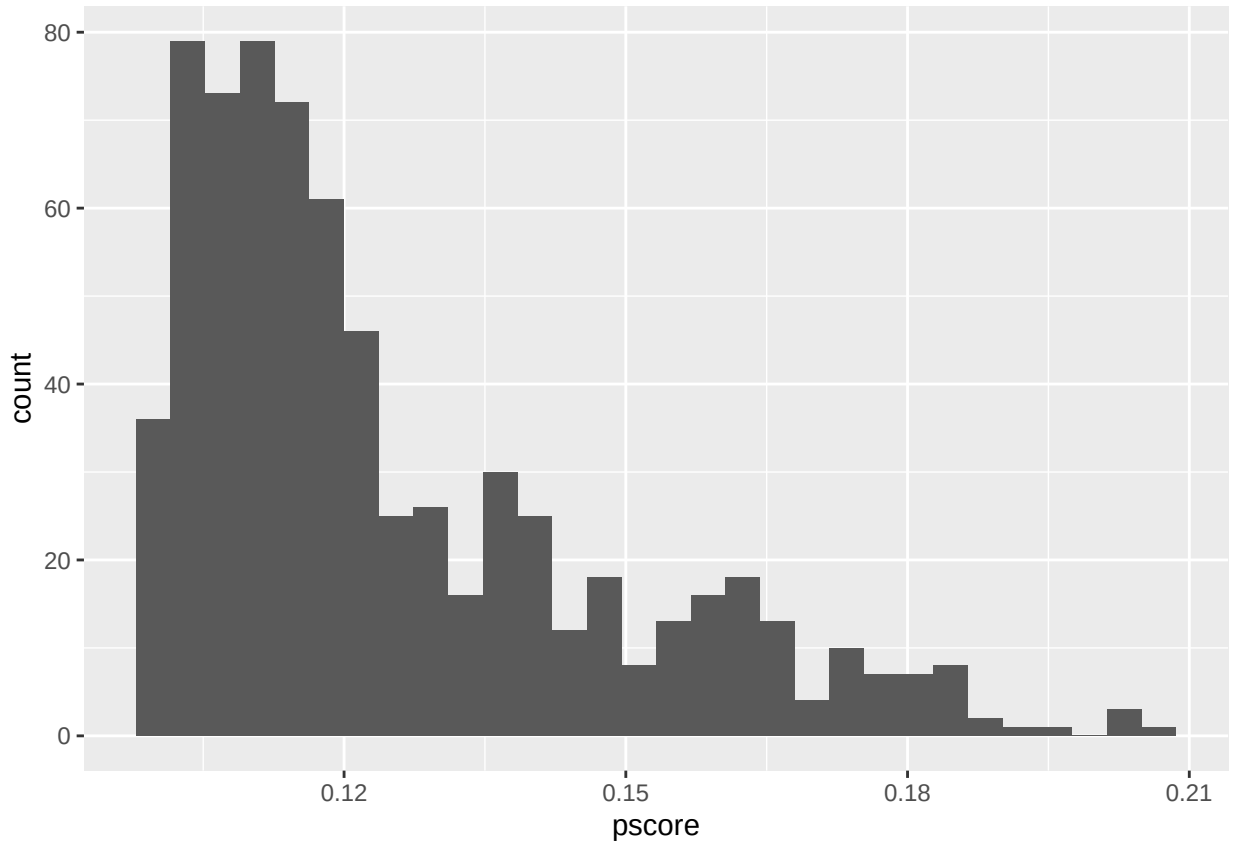

```
# linear probability model (quadratic) treatment
nsw_OLS_quad_trim %>%
  filter(treat == 1) %>%
  ggplot() +
  geom_histogram(aes(x = pscore))
```

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



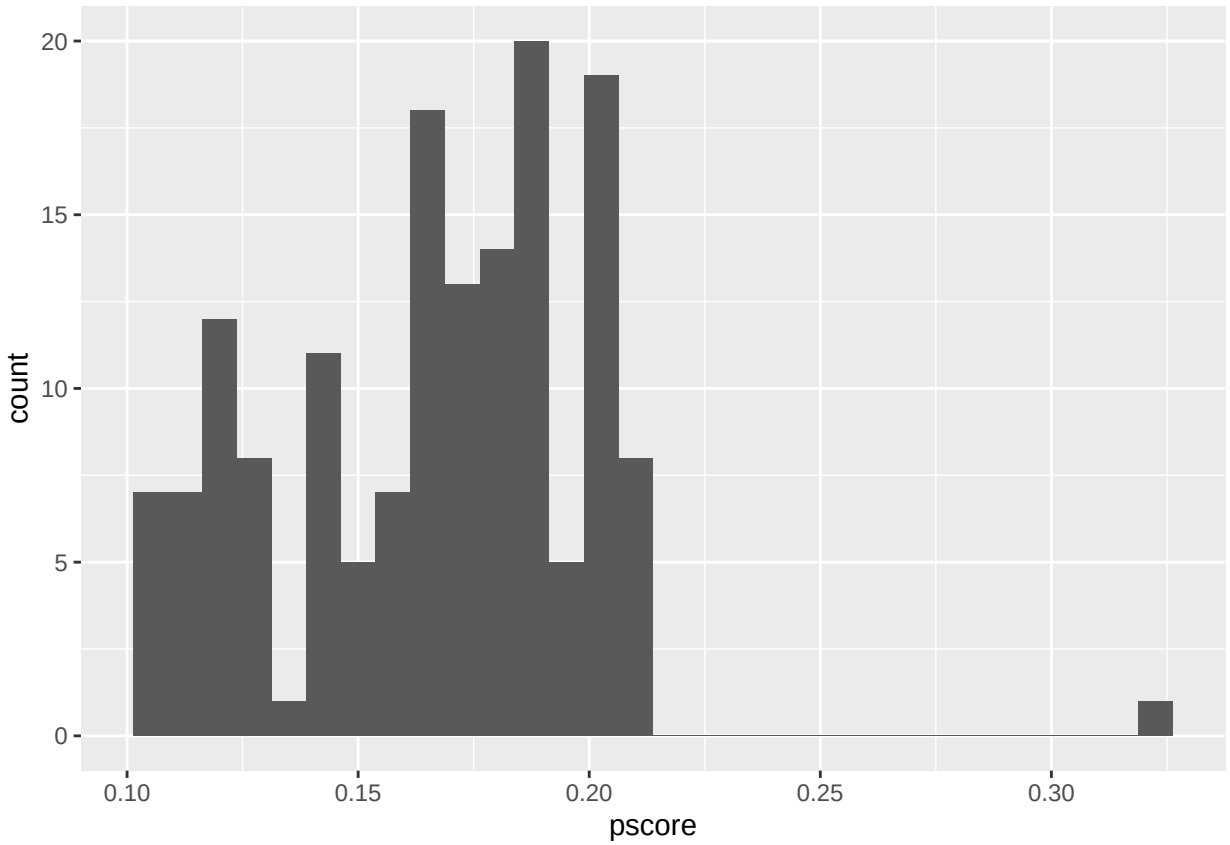

```
# linear probability model (cubic) control
nsw_OLS_cube_trim %>%
  filter(treat == 0) %>%
  ggplot() +
  geom_histogram(aes(x = pscore))
```

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




```
# linear probability model (cubic) treatment
nsw_OLS_cube_trim %>%
  filter(treat == 1) %>%
  ggplot() +
  geom_histogram(aes(x = pscore))
```

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



For the following questions I will only use the trimmed data, cubic logit model, and the quadratic linear probability model.

Question 2

```
##### Question 2: First difference for the logit cubic model and quadratic linear probability model
### First difference for the logit cubic model
# Filter re78 to only include people in the treatment group
nsw_logit_cube_trim %>%
  filter(treat == 1) %>%
  summary(re78)
```

```
##    data_id              treat        age              educ           black
##  Length:138         Min.   :1   Min.   :17.00   Min.   : 4.00   Min.   :0.000
##  Class :character   1st Qu.:1   1st Qu.:20.00   1st Qu.: 9.00   1st Qu.:1.000
##  Mode  :character   Median :1   Median :25.00   Median :11.00   Median :1.000
##                     Mean   :1   Mean   :25.27   Mean   :10.26   Mean   :0.942
##                     3rd Qu.:1   3rd Qu.:28.00   3rd Qu.:11.00   3rd Qu.:1.000
##                     Max.   :1   Max.   :48.00   Max.   :15.00   Max.   :1.000
##       hisp              marr           nodegree           re74
##  Min.   :0.00000   Min.   :0.0000   Min.   :0.0000   Min.   :    0
##  1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:1.0000   1st Qu.:    0
##  Median :0.00000   Median :0.0000   Median :1.0000   Median :    0
##  Mean   :0.03623   Mean   :0.1449   Mean   :0.7536   Mean   : 1788
##  3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:    0
##  Max.   :1.00000   Max.   :1.0000   Max.   :1.0000   Max.   :35040
##       re75            re78           agesq           agecube
```

```
##  Min.   :    0   Min.   :    0   Min.   : 289.0   Min.   :  4913
##  1st Qu.:    0   1st Qu.:    0   1st Qu.: 400.0   1st Qu.:  8000
##  Median :    0   Median : 4003   Median : 625.0   Median : 15625
##  Mean   : 1084   Mean   : 6095   Mean   : 682.3   Mean   : 19799
##  3rd Qu.: 1621   3rd Qu.: 9589   3rd Qu.: 784.0   3rd Qu.: 21952
##  Max.   :11537   Max.   :36647   Max.   :2304.0   Max.   :110592
##      educsq         educcube         u74              u75
##  Min.   : 16.0   Min.   :  64   Min.   :0.0000   Min.   :0.0000
##  1st Qu.: 81.0   1st Qu.: 729   1st Qu.:1.0000   1st Qu.:0.0000
##  Median :121.0   Median :1331   Median :1.0000   Median :1.0000
##  Mean   :109.1   Mean   :1190   Mean   :0.7609   Mean   :0.6449
##  3rd Qu.:121.0   3rd Qu.:1331   3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :225.0   Max.   :3375   Max.   :1.0000   Max.   :1.0000
##      re74sq             re74cube             re75sq
##  Min.   :0.000e+00   Min.   :0.000e+00   Min.   :        0
##  1st Qu.:0.000e+00   1st Qu.:0.000e+00   1st Qu.:        0
##  Median :0.000e+00   Median :0.000e+00   Median :        0
##  Mean   :2.633e+07   Mean   :5.813e+11   Mean   :  5279631
##  3rd Qu.:0.000e+00   3rd Qu.:0.000e+00   3rd Qu.:  2633251
##  Max.   :1.228e+09   Max.   :4.302e+13   Max.   :133092455
##     re75cube             re78sq           interaction1     interaction2
##  Min.   :0.000e+00   Min.   :0.000e+00   Min.   :     0   Min.   :0.00000
##  1st Qu.:0.000e+00   1st Qu.:0.000e+00   1st Qu.:     0   1st Qu.:0.00000
##  Median :6.414e+10   Median :1.602e+07   Median :     0   Median :0.00000
##  Mean   :1.818e+12   Mean   :8.909e+07   Mean   : 19684   Mean   :0.03623
##  3rd Qu.:8.816e+11   3rd Qu.:9.194e+07   3rd Qu.:     0   3rd Qu.:0.00000
##  Max.   :4.922e+13   Max.   :1.343e+09   Max.   :490561   Max.   :1.00000
##      pscore
##  Min.   :0.1031
##  1st Qu.:0.3040
##  Median :0.5097
##  Mean   :0.5035
##  3rd Qu.:0.6873
##  Max.   :0.8985
```

```
# Filter and average on the treatment effect
mean1 <- nsw_logit_cube_trim %>%
  filter(treat == 1) %>%
  pull(re78) %>%
  mean()

# save the
nsw_logit_cube_trim$y1 <- mean1

# Average treatment effect on the untreated
nsw_logit_cube_trim %>%
  filter(treat == 0) %>%
  summary(re78)
```

```
##    data_id              treat        age             educ          black
##  Length:224         Min.   :0   Min.   :16.00   Min.   : 2.00   Min.   :0.000
##  Class :character   1st Qu.:0   1st Qu.:19.00   1st Qu.: 9.00   1st Qu.:1.000
##  Mode  :character   Median :0   Median :24.00   Median :11.00   Median :1.000
##                     Mean   :0   Mean   :26.08   Mean   :10.38   Mean   :0.942
##                     3rd Qu.:0   3rd Qu.:31.00   3rd Qu.:12.00   3rd Qu.:1.000
```

```
##                    Max.   :0   Max.   :55.00   Max.   :16.00   Max.   :1.000
##       hisp              marr           nodegree           re74
##  Min.   :0.00000   Min.   :0.0000   Min.   :0.0000   Min.   :    0
##  1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:    0
##  Median :0.00000   Median :0.0000   Median :1.0000   Median :    0
##  Mean   :0.03571   Mean   :0.2054   Mean   :0.6295   Mean   : 2426
##  3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.: 3511
##  Max.   :1.00000   Max.   :1.0000   Max.   :1.0000   Max.   :22322
##       re75              re78           agesq           agecube
##  Min.   :    0.00   Min.   :    0   Min.   : 256.0   Min.   :  4096
##  1st Qu.:    0.00   1st Qu.:    0   1st Qu.: 361.0   1st Qu.:  6859
##  Median :   50.13   Median : 1632   Median : 576.0   Median : 13824
##  Mean   : 1630.70   Mean   : 4633   Mean   : 748.2   Mean   : 23629
##  3rd Qu.: 2258.04   3rd Qu.: 7696   3rd Qu.: 961.0   3rd Qu.: 29791
##  Max.   :13117.69   Max.   :25565   Max.   :3025.0   Max.   :166375
##      educsq          educcube          u74              u75
##  Min.   :  4.0   Min.   :   8   Min.   :0.0000   Min.   :0.0000
##  1st Qu.: 81.0   1st Qu.: 729   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :121.0   Median :1331   Median :1.0000   Median :0.0000
##  Mean   :112.7   Mean   :1264   Mean   :0.5759   Mean   :0.4911
##  3rd Qu.:144.0   3rd Qu.:1728   3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :256.0   Max.   :4096   Max.   :1.0000   Max.   :1.0000
##      re74sq             re74cube             re75sq
##  Min.   :        0   Min.   :0.000e+00   Min.   :        0
##  1st Qu.:        0   1st Qu.:0.000e+00   1st Qu.:        0
##  Median :        0   Median :0.000e+00   Median :     2542
##  Mean   : 23138265   Mean   :2.921e+11   Mean   : 10588462
##  3rd Qu.: 12345738   3rd Qu.:4.346e+10   3rd Qu.:  5098777
##  Max.   :498268545   Max.   :1.112e+13   Max.   :172073802
##     re75cube             re78sq           interaction1     interaction2
##  Min.   :0.000e+00   Min.   :        0   Min.   :     0   Min.   :0.00000
##  1st Qu.:0.000e+00   1st Qu.:        0   1st Qu.:     0   1st Qu.:0.00000
##  Median :4.395e+09   Median :  2673567   Median :     0   Median :0.00000
##  Mean   :8.944e+11   Mean   : 57217111   Mean   : 23727   Mean   :0.02232
##  3rd Qu.:4.559e+11   3rd Qu.: 59229890   3rd Qu.: 37942   3rd Qu.:0.00000
##  Max.   :1.671e+13   Max.   :653552348   Max.   :267863   Max.   :1.00000
##      pscore             y1
##  Min.   :0.1001   Min.   :6095
##  1st Qu.:0.1454   1st Qu.:6095
##  Median :0.2300   Median :6095
##  Mean   :0.2893   Mean   :6095
##  3rd Qu.:0.3766   3rd Qu.:6095
##  Max.   :0.8935   Max.   :6095
```

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

nsw_logit_cube_trim$y0 <- mean0

ate_logit <- unique(nsw_logit_cube_trim$y1 - nsw_logit_cube_trim$y0)
# 1461.0585
nsw_logit_cube_trim <- nsw_logit_cube_trim %>%
```

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

### First difference for quadratic linear probability model
# Filter re78 to only include people in the treatment group
nsw_OLS_quad_trim %>%
  filter(treat == 1) %>%
  summary(re78)
```

```
##    data_id              treat        age             educ           black
##  Length:152         Min.   :1   Min.   :17.00   Min.   : 4.00   Min.   :1
##  Class :character   1st Qu.:1   1st Qu.:20.00   1st Qu.: 9.00   1st Qu.:1
##  Mode  :character   Median :1   Median :25.00   Median :11.00   Median :1
##                     Mean   :1   Mean   :25.93   Mean   :10.26   Mean   :1
##                     3rd Qu.:1   3rd Qu.:29.00   3rd Qu.:12.00   3rd Qu.:1
##                     Max.   :1   Max.   :48.00   Max.   :16.00   Max.   :1
##       hisp        marr           nodegree           re74
##  Min.   :0   Min.   :0.0000   Min.   :0.0000   Min.   :    0.0
##  1st Qu.:0   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:    0.0
##  Median :0   Median :0.0000   Median :1.0000   Median :    0.0
##  Mean   :0   Mean   :0.1645   Mean   :0.7434   Mean   : 1768.3
##  3rd Qu.:0   3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:  989.3
##  Max.   :0   Max.   :1.0000   Max.   :1.0000   Max.   :25929.7
##       re75            re78           agesq           agecube
##  Min.   :    0   Min.   :    0   Min.   : 289.0   Min.   :  4913
##  1st Qu.:    0   1st Qu.:    0   1st Qu.: 400.0   1st Qu.:  8000
##  Median :    0   Median : 3880   Median : 625.0   Median : 15625
##  Mean   : 1342   Mean   : 5942   Mean   : 726.6   Mean   : 22086
##  3rd Qu.: 1674   3rd Qu.: 9292   3rd Qu.: 841.0   3rd Qu.: 24389
##  Max.   :25142   Max.   :60308   Max.   :2304.0   Max.   :110592
##      educsq         educcube         u74              u75
##  Min.   : 16.0   Min.   :  64   Min.   :0.0000   Min.   :0.0000
##  1st Qu.: 81.0   1st Qu.: 729   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :121.0   Median :1331   Median :1.0000   Median :1.0000
##  Mean   :109.4   Mean   :1201   Mean   :0.7303   Mean   :0.6316
##  3rd Qu.:144.0   3rd Qu.:1728   3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :256.0   Max.   :4096   Max.   :1.0000   Max.   :1.0000
##      re74sq             re74cube             re75sq
##  Min.   :        0   Min.   :0.000e+00   Min.   :        0
##  1st Qu.:        0   1st Qu.:0.000e+00   1st Qu.:        0
##  Median :        0   Median :0.000e+00   Median :        0
##  Mean   : 20091417   Mean   :3.012e+11   Mean   : 11477633
##  3rd Qu.:   978651   3rd Qu.:9.681e+08   3rd Qu.:  2803266
##  Max.   :672348289   Max.   :1.743e+13   Max.   :632132244
##     re75cube             re78sq           interaction1     interaction2
##  Min.   :0.000e+00   Min.   :0.000e+00   Min.   :     0   Min.   :0
##  1st Qu.:0.000e+00   1st Qu.:0.000e+00   1st Qu.:     0   1st Qu.:0
##  Median :5.849e+10   Median :1.506e+07   Median :     0   Median :0
##  Mean   :2.700e+12   Mean   :9.575e+07   Mean   : 18271   Mean   :0
##  3rd Qu.:8.024e+11   3rd Qu.:8.635e+07   3rd Qu.:  9583   3rd Qu.:0
##  Max.   :2.193e+14   Max.   :3.637e+09   Max.   :259297   Max.   :0
##      pscore
##  Min.   :0.1018
##  1st Qu.:0.1310
```

```
##  Median :0.1709
##  Mean   :0.1598
##  3rd Qu.:0.1793
##  Max.   :0.1935
```

```
# Filter and average on the treatment effect
mean1_OLS <- nsw_OLS_quad_trim %>%
  filter(treat == 1) %>%
  pull(re78) %>%
  mean()

# save the
nsw_OLS_quad_trim$y1_OLS <- mean1_OLS

# Average treatment effect on the untreated
nsw_OLS_quad_trim %>%
  filter(treat == 0) %>%
  summary(re78)
```

```
##    data_id              treat        age              educ           black
##  Length:769         Min.   :0   Min.   :16.00   Min.   : 0.00   Min.   :1
##  Class :character   1st Qu.:0   1st Qu.:22.00   1st Qu.: 9.00   1st Qu.:1
##  Mode  :character   Median :0   Median :29.00   Median :11.00   Median :1
##                     Mean   :0   Mean   :30.72   Mean   :10.89   Mean   :1
##                     3rd Qu.:0   3rd Qu.:38.00   3rd Qu.:12.00   3rd Qu.:1
##                     Max.   :0   Max.   :55.00   Max.   :18.00   Max.   :1
##       hisp        marr           nodegree           re74
##  Min.   :0   Min.   :0.0000   Min.   :0.0000   Min.   :    0.0
##  1st Qu.:0   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:  525.1
##  Median :0   Median :0.0000   Median :1.0000   Median : 8060.4
##  Mean   :0   Mean   :0.4499   Mean   :0.5267   Mean   : 9965.0
##  3rd Qu.:0   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:17827.4
##  Max.   :0   Max.   :1.0000   Max.   :1.0000   Max.   :25862.3
##       re75             re78             agesq         agecube
##  Min.   :    0.0   Min.   :    0.0   Min.   : 256   Min.   :  4096
##  1st Qu.:  136.1   1st Qu.:  731.5   1st Qu.: 484   1st Qu.: 10648
##  Median : 6939.3   Median : 9339.2   Median : 841   Median : 24389
##  Mean   : 9309.3   Mean   :10956.5   Mean   :1049   Mean   : 39104
##  3rd Qu.:17135.2   3rd Qu.:19265.1   3rd Qu.:1444   3rd Qu.: 54872
##  Max.   :25243.6   Max.   :25564.7   Max.   :3025   Max.   :166375
##      educsq          educcube         u74              u75
##  Min.   :  0.0   Min.   :   0   Min.   :0.0000   Min.   :0.000
##  1st Qu.: 81.0   1st Qu.: 729   1st Qu.:0.0000   1st Qu.:0.000
##  Median :121.0   Median :1331   Median :0.0000   Median :0.000
##  Mean   :125.3   Mean   :1504   Mean   :0.2146   Mean   :0.238
##  3rd Qu.:144.0   3rd Qu.:1728   3rd Qu.:0.0000   3rd Qu.:0.000
##  Max.   :324.0   Max.   :5832   Max.   :1.0000   Max.   :1.000
##      re74sq             re74cube            re75sq
##  Min.   :        0   Min.   :0.000e+00   Min.   :        0
##  1st Qu.:   275713   1st Qu.:1.448e+08   1st Qu.:    18514
##  Median : 64970432   Median :5.237e+11   Median : 48153746
##  Mean   :186586136   Mean   :3.972e+12   Mean   :169529862
##  3rd Qu.:317815090   3rd Qu.:5.666e+12   3rd Qu.:293614383
##  Max.   :668859612   Max.   :1.730e+13   Max.   :637236856
##     re75cube             re78sq           interaction1     interaction2
```

```
##  Min.   :0.000e+00   Min.   :        0   Min.   :     0   Min.   :0
##  1st Qu.:3.914e+08   1st Qu.:   535055   1st Qu.:  5167   1st Qu.:0
##  Median :8.146e+11   Median : 87221262   Median : 82113   Median :0
##  Mean   :4.507e+12   Mean   :209689105   Mean   :108583   Mean   :0
##  3rd Qu.:7.150e+12   3rd Qu.:371144816   3rd Qu.:174590   3rd Qu.:0
##  Max.   :1.671e+13   Max.   :653552348   Max.   :439659   Max.   :0
##      pscore           y1_OLS
##  Min.   :0.1000   Min.   :5942
##  1st Qu.:0.1065   1st Qu.:5942
##  Median :0.1137   Median :5942
##  Mean   :0.1225   Mean   :5942
##  3rd Qu.:0.1313   3rd Qu.:5942
##  Max.   :0.1935   Max.   :5942
```

```
mean0_OLS <- nsw_OLS_quad_trim %>%
  filter(treat == 0) %>%
  pull(re78) %>%
  mean()

nsw_OLS_quad_trim$y0_OLS <- mean0_OLS

ate_OLS <- unique(nsw_OLS_quad_trim$y1_OLS - nsw_OLS_quad_trim$y0_OLS)
# -5014.1128
nsw_OLS_quad_trim <- nsw_OLS_quad_trim %>%
  filter(treat == 1) %>%
  select(-y1_OLS, -y0_OLS)
```

The average treatment effect for the cubic logit model is 1461.06 and the average treatment effect for the quadratic linear probability model is -5014.11.

Question 3

```
##### Question 3 weighted difference in difference regression
### Cubic logit model
N <- nrow(nsw_logit_cube_trim)
# Manual with non-normalized weights using trimmed data for the cubic logit model
nsw_logit_cube_trim <- nsw_logit_cube_trim %>%
  mutate(d1 = treat/pscore,
         d0 = (1-treat)/(1-pscore))

s1 <- sum(nsw_logit_cube_trim$d1)
s0 <- sum(nsw_logit_cube_trim$d0)

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

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

```
## [1] 16029.51
```

The weighted difference in difference estimate for the cubic logit model is 16029.51.

```r
### Quadratic linear proability model
# Manual with non-normalized weights using trimmed data for the cubic logit model
nsw_OLS_quad_trim <- nsw_OLS_quad_trim %>%
  mutate(d1 = treat/pscore,
         d0 = (1-treat)/(1-pscore))

s1 <- sum(nsw_OLS_quad_trim$d1)
s0 <- sum(nsw_OLS_quad_trim$d0)

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

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

```
## [1] 38049.65
```

The weighted difference in difference estimate for the quadratic linear proability model is -4047.76.