

Propensity Score Analysis

EPSY 630 - Statistics II

Jason Bryer, Ph.D.

April 13, 2021

Agenda

- Example Data Project
- Propensity Score Analysis
- One Minute Papers

Data Project

Checklist / Suggested Outline

- Overview slide
 - Context on the data collection
 - Description of the dependent variable (what is being measured)
 - Description of the independent variable (what is being measured; include at least 2 variables)
 - Research question
- Summary statistics
- Include appropriate data visualizations.
- Statistical output
 - Include the appropriate statistics for your method used.
 - For null hypothesis tests (e.g. t-test, chi-squared, ANOVA, etc.), state the null and alternative hypotheses along with relevant statistic and p-value (and confidence interval if appropriate).
 - For regression models, include the regression output and interpret the R-squared value.
- Conclusion
 - Why is this analysis important?
 - Limitations of the analysis?

Criteria for Grading

- Data is presented to support the conclusions using the appropriate analysis (i.e. the statistical method chosen supports the research question).
- Suitable tables summarize data in a clear and meaningful way even to those unfamiliar with the project.
- Suitable graphics summarize data in a clear and meaningful way even to those unfamiliar with the project.
- Data reviewed and analyzed accurately and coherently.
- Proper use of descriptive and/or inferential statistics.

Example Project

2000 Election

The 2000 election between George Bush and Al Gore was ultimately decided in Florida. However, there was a third candidate on the ballot, Pat Buchanan, and one county with an unpredictable outcome. Is there evidence that a large number of votes were cast for a mistaken candidate?

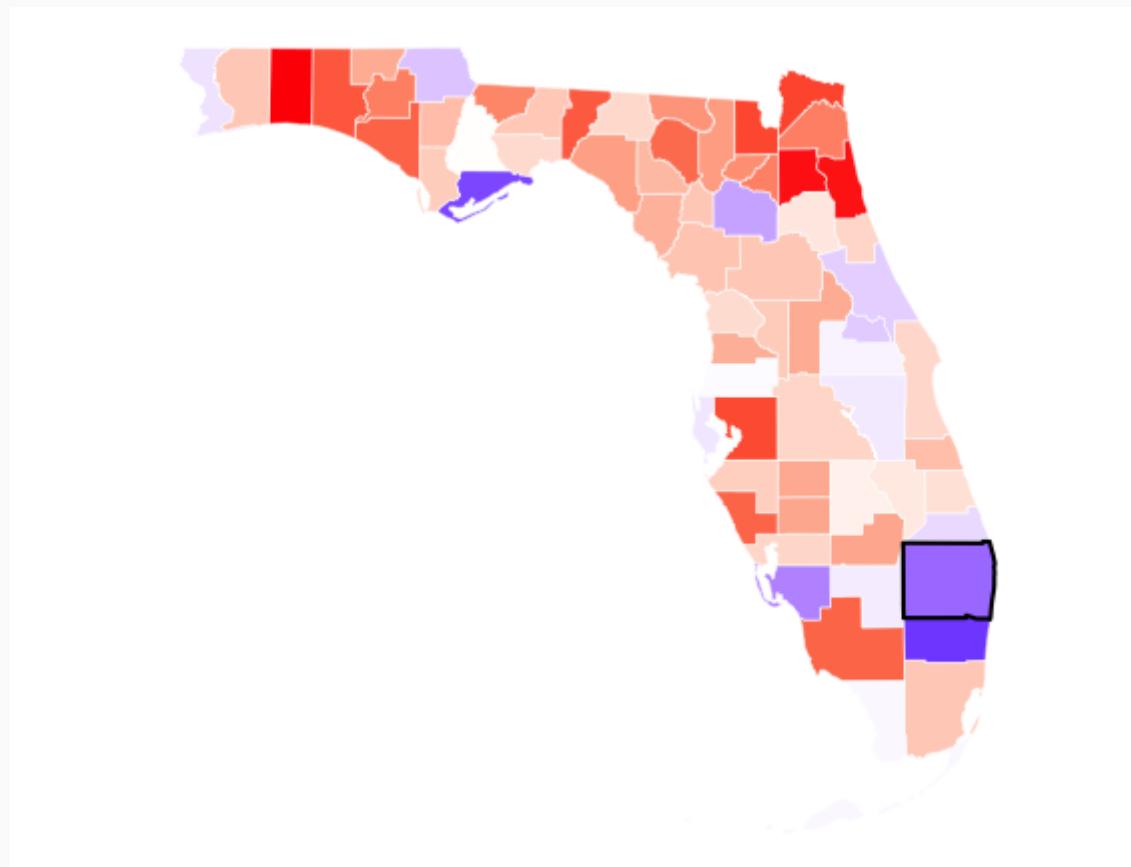
The `elections` data frame contains the breakdown of votes by each of the 67 counties in Florida.

```
elections <- read.table("../course_data/2000elections.txt", header=TRUE)
```

There are 67 counties in Florida that cast a total of 2,910,078 votes for George Bush and 2,909,117 resulting in Bush winning by 961 votes.

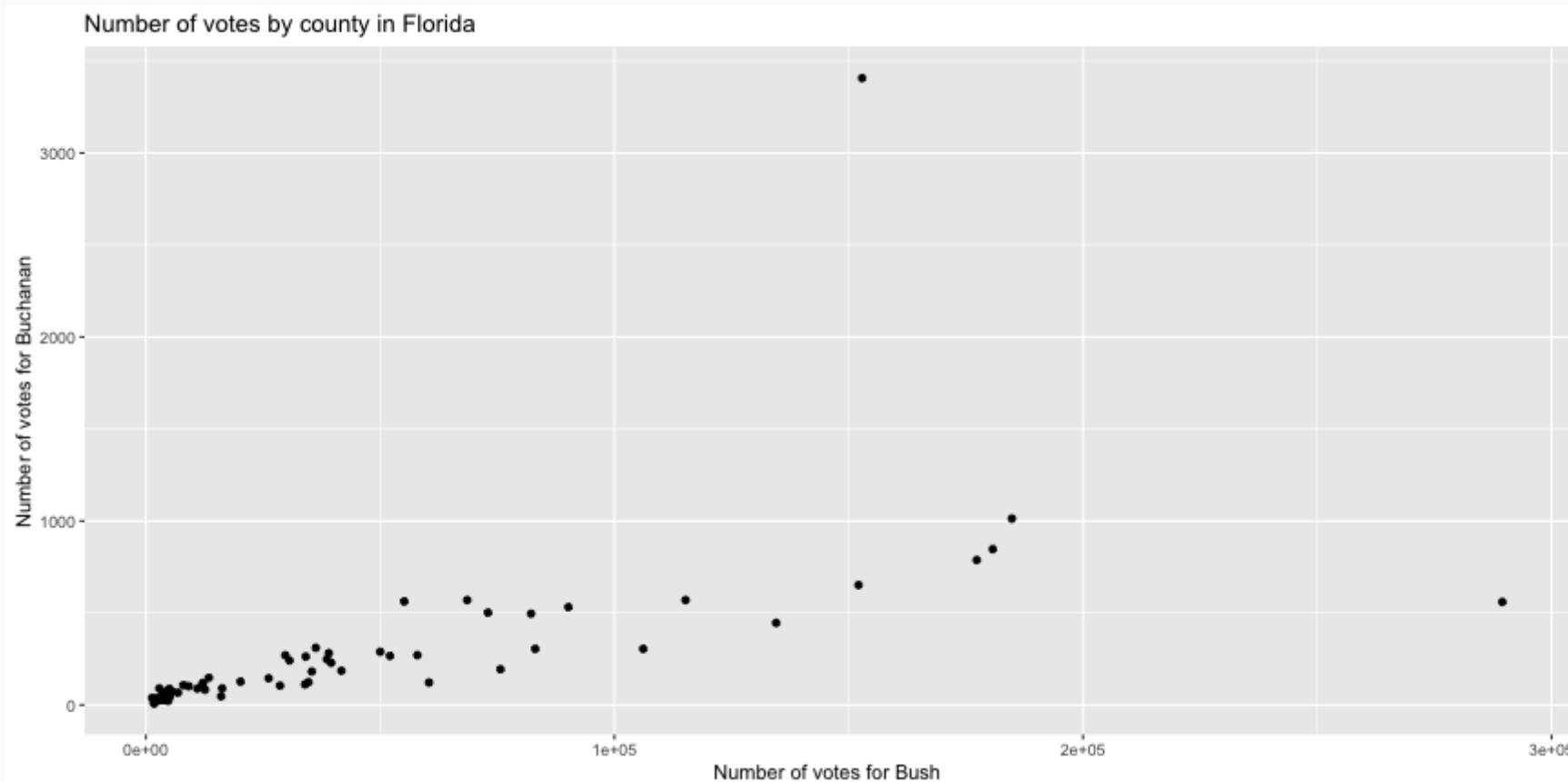
However, in the days following the election there was much controversy surrounding so called "hanging chads." That is, there were a number of ballots where it was not clear who the vote was for. This was a particular issue in Palm Beach.

Florida Counties (blue = Gore; red = Bush)



Number of votes by county in Florida

```
ggplot(elections, aes(bush, buch)) + geom_point() +  
  xlab("Number of votes for Bush") + ylab("Number of votes for Buchanan") +  
  ggtitle("Number of votes by county in Florida")
```



Correlation

```
cor.test(elections$buch, elections$bush)
```

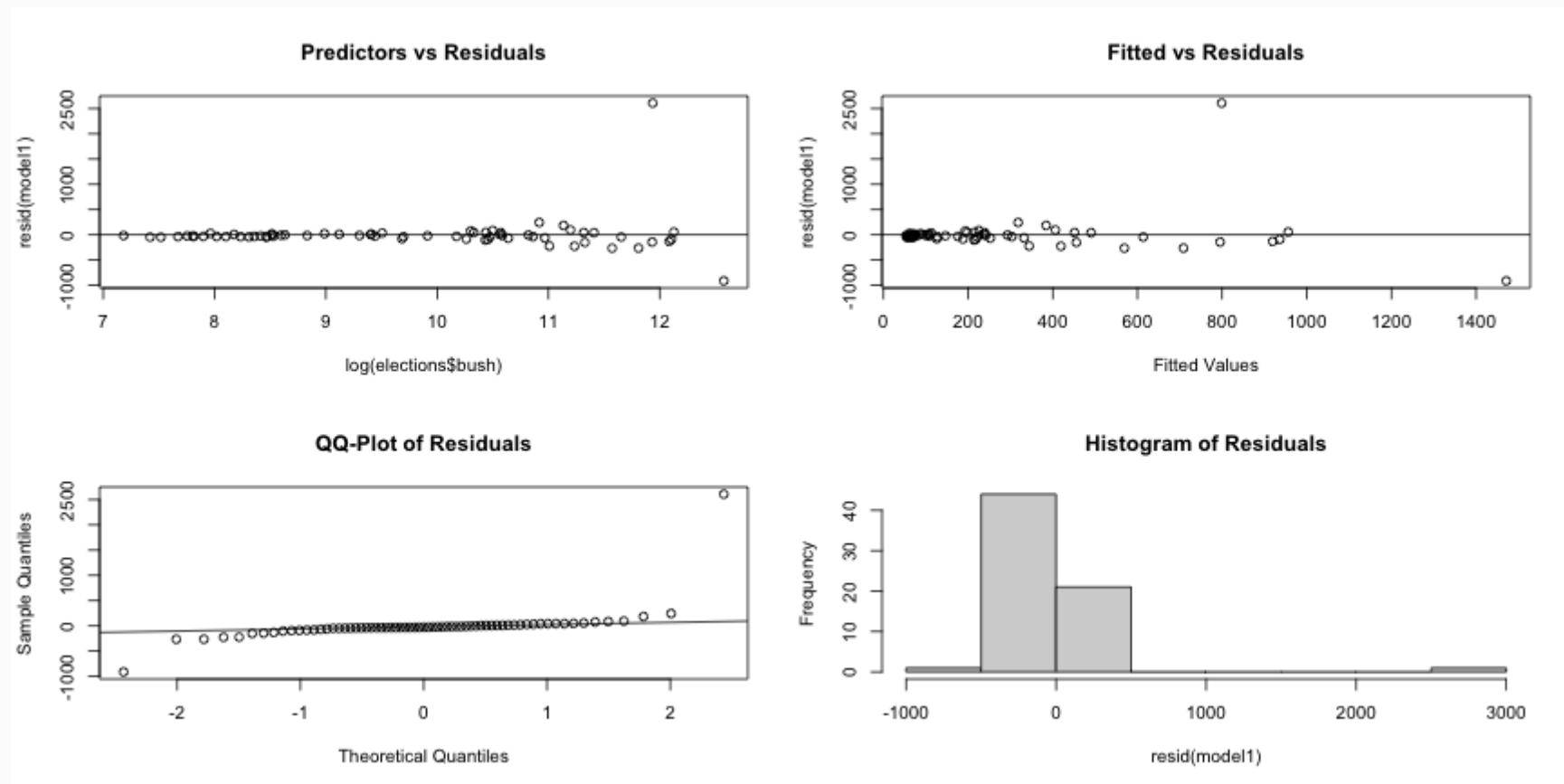
```
##  
##      Pearson's product-moment correlation  
##  
## data: elections$buch and elections$bush  
## t = 6.455, df = 65, p-value = 1.574e-08  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
##  0.4527668 0.7522709  
## sample estimates:  
##      cor  
## 0.6250012
```

Linear Regression Model

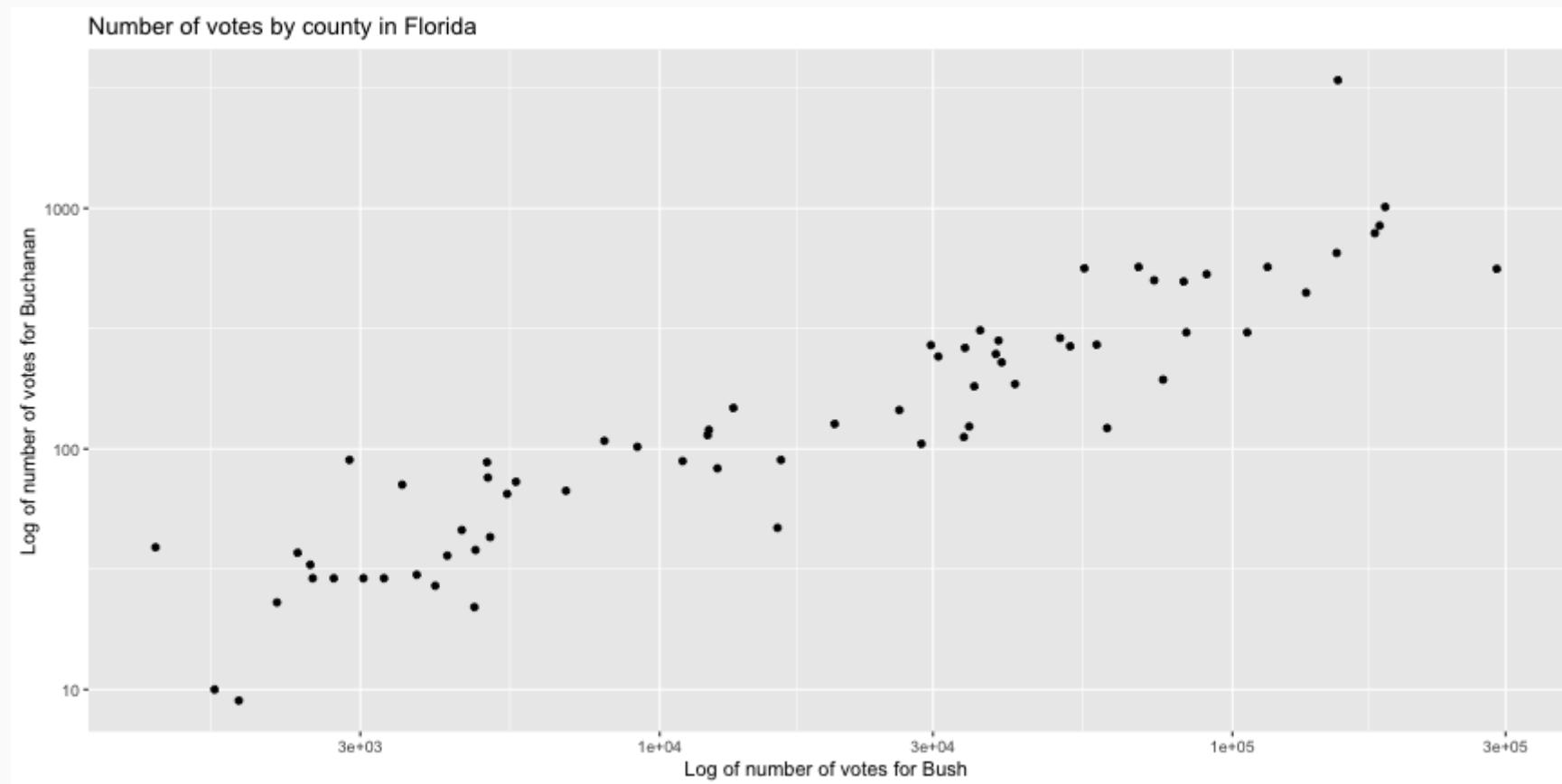
```
model1 <- lm(buch ~ bush, data = elections)
summary(model1)
```

```
##
## Call:
## lm(formula = buch ~ bush, data = elections)
##
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -911.30  -46.11  -26.05   12.01 2608.01 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 4.697e+01  5.446e+01   0.863   0.392    
## bush        4.920e-03  7.622e-04   6.455 1.57e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 353.9 on 65 degrees of freedom
## Multiple R-squared:  0.3906,    Adjusted R-squared:  0.3813 
## F-statistic: 41.67 on 1 and 65 DF,  p-value: 1.574e-08
```

Residual Analysis



Log Tranform



Correlation with log tranformations

```
cor.test(log(elections$buch), log(elections$bush))
```

```
##  
##      Pearson's product-moment correlation  
  
## data: log(elections$buch) and log(elections$bush)  
## t = 19.222, df = 65, p-value < 2.2e-16  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
##  0.8760098 0.9515894  
## sample estimates:  
##      cor  
## 0.9221706
```

Linear Regression Model (log transform)

```
model2 <- lm(log(buch) ~ log(bush), data = elections)
summary(model2)
```

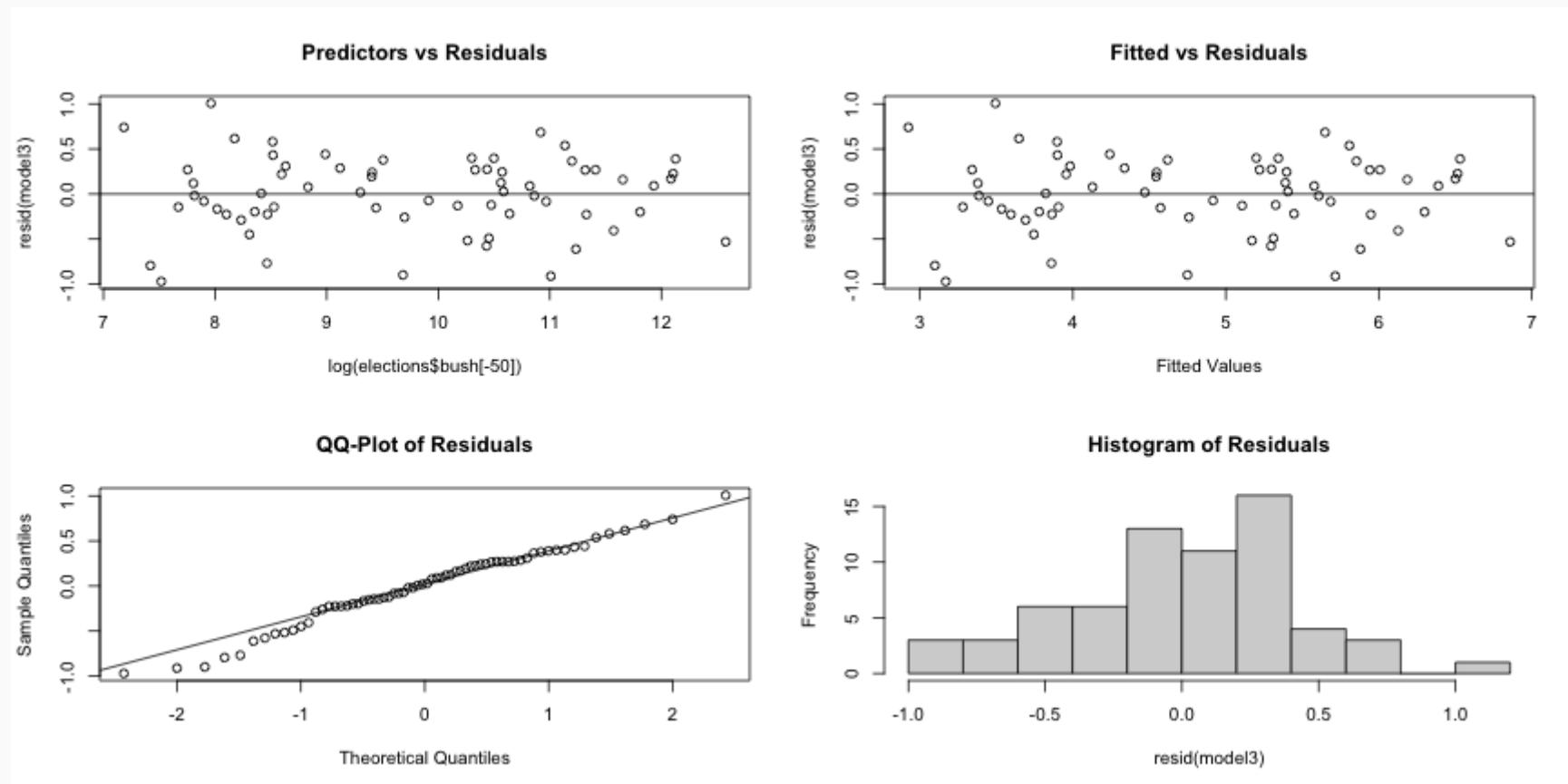
```
##
## Call:
## lm(formula = log(buch) ~ log(bush), data = elections)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.97038 -0.24247  0.00825  0.25452  1.65752
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.55079    0.38903 -6.557 1.04e-08 ***
## log(bush)    0.75620    0.03934 19.222 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4672 on 65 degrees of freedom
## Multiple R-squared:  0.8504,    Adjusted R-squared:  0.8481
## F-statistic: 369.5 on 1 and 65 DF,  p-value: < 2.2e-16
```

Regression model without Palm Beach

```
model3 <- lm(log(buch) ~ log(bush), data = elections[-50,])
summary(model3)
```

```
## 
## Call:
## lm(formula = log(buch) ~ log(bush), data = elections[-50, ])
## 
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -0.97136 -0.22384  0.02279  0.26959  1.00652 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) -2.31657    0.35470  -6.531 1.23e-08 ***
## log(bush)     0.72960    0.03599   20.271 < 2e-16 ***
## ---      
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 
## 
## Residual standard error: 0.4203 on 64 degrees of freedom
## Multiple R-squared:  0.8652,    Adjusted R-squared:  0.8631 
## F-statistic: 410.9 on 1 and 64 DF,  p-value: < 2.2e-16
```

Residual Analysis (log)



Predict Palm Beach from the model

Obtain the predicted vote count for Palm Beach given the fitted model without

```
new <- data.frame(bush = elections$bush[50])
```

The difference between predicted on the original scale and the observed vote count

```
elections$buch[50] - exp(predict(model3, new))
```

```
##      1  
## 2809.498
```

Predict Palm Beach from the model (cont.)

Prediction Confidence Interval for log(vote count)

```
predict(model3, new, interval='prediction', level=.95)
```

```
##      fit     lwr     upr
## 1 6.392757 5.532353 7.253162
```

Prediction Confidence Interval on the original scale

```
exp(predict(model3, new, interval='prediction', level=.95))
```

```
##      fit     lwr     upr
## 1 597.5019 252.738 1412.564
```

```
elections$buch[50]
```

```
## [1] 3407
```

Therefore, what we can say is that it is likely that Palm Beach is a different community.

References

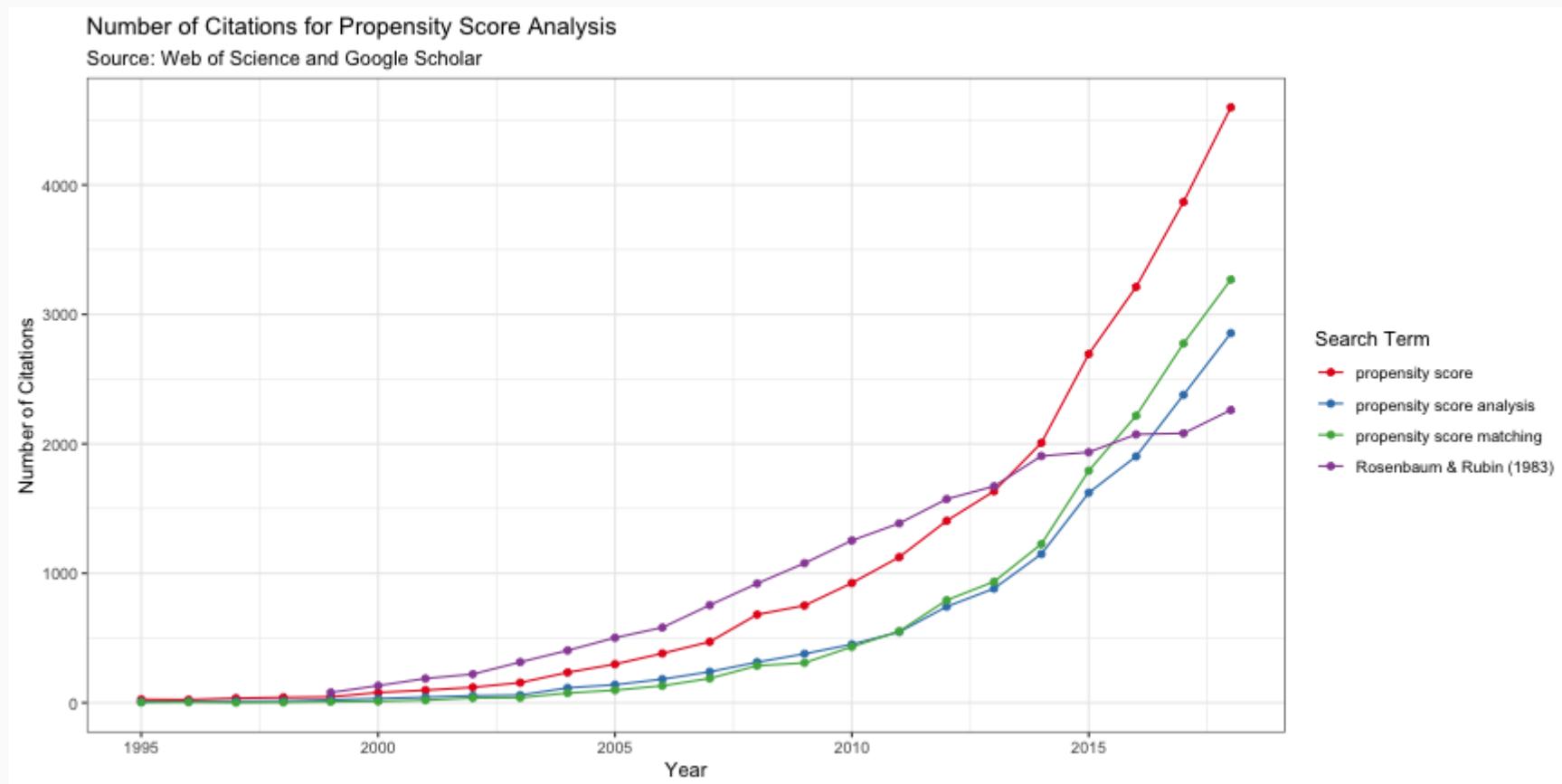
Wand, J.N., Shotts, K.W., Sekhon, J.S., Mebane, W.R., Herron, M.C., & Brady, H.E. (2001). *The Butterfly Did It: The Aberrant Vote for Buchanan in Palm Beach County, Florida*. *American Political Science Review*, 95(4).

Smith, R.L. (2002). *A Statistical Assessment of Buchanan's Vote in Palm Beach County*. *Statistical Science*, 17(4).

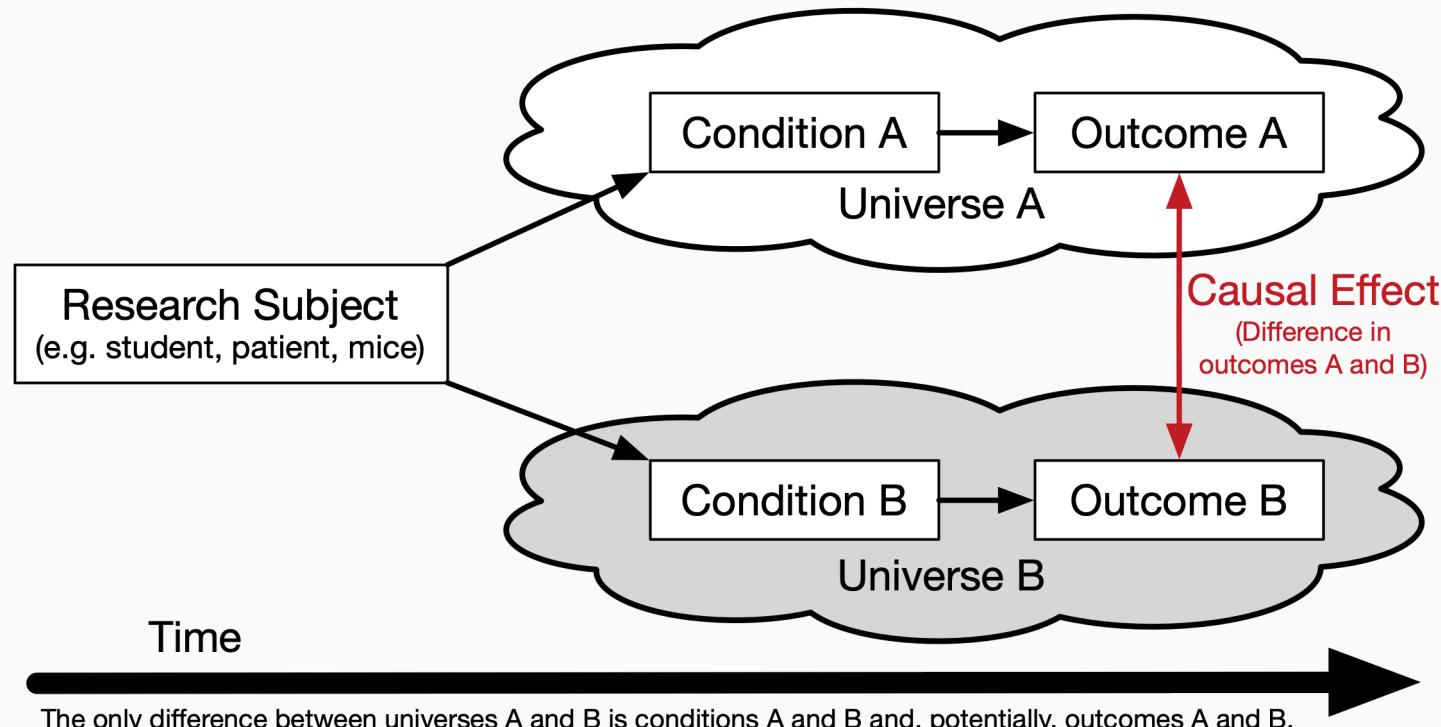
Regression Analysis - 1.3.3 - 2000 Elections in Florida (Youtube)

Propensity Score Analysis

Popularity of Propensity Score Analysis



Counterfactuals



The Randomized Experiment

Considered to be the *gold standard* for estimating causal effects.

- Effects can be estimated using simple means between groups, or blocks in randomized block design.
- Randomization presumes unbiasedness and balance between groups.

However, randomization is often not feasible for many reasons, especially in educational contexts.

The strong ignorability assumption states that:

$$(Y_i(1), Y_i(0)) \perp\!\!\!\perp T_i | X_i = x$$

for all X_i .

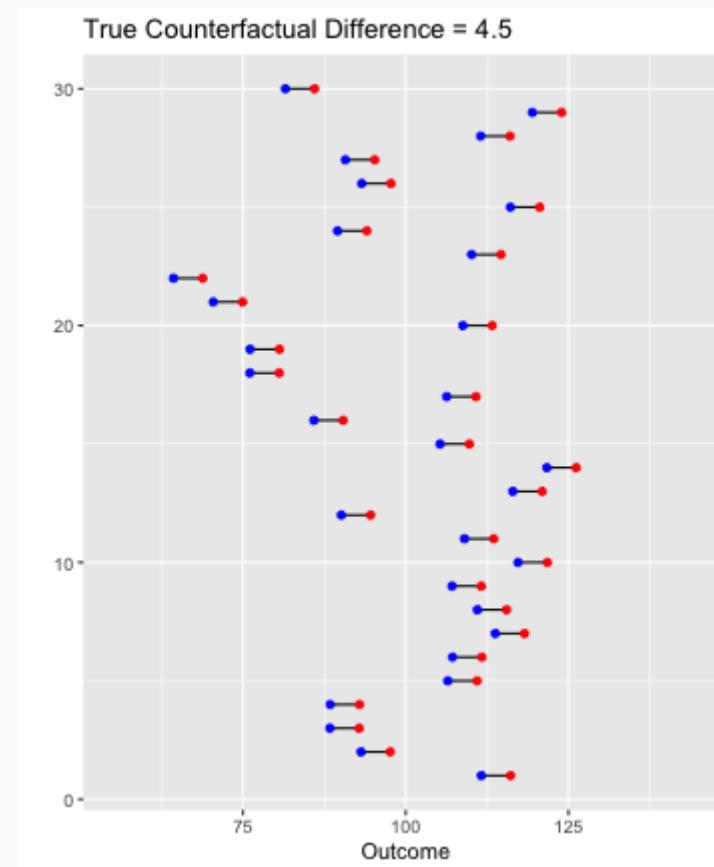
RCT Example

```
set.seed(2112)
pop.mean <- 100
pop.sd <- 15
pop.es <- .3
n <- 30
thedata <- data.frame(id = 1:30, center = rnorm(n, mean = pop.mean, sd = pop.sd))
)
val <- pop.sd * pop.es / 2
thedata$placebo <- thedata$center - val
thedata$treatment <- thedata$center + val
thedata$diff <- thedata$treatment - thedata$placebo
thedata$RCT_Assignment <- sample(c('placebo', 'treatment'), n, replace = TRUE)
thedata$RCT_Value <- as.numeric(apply(thedata, 1, FUN = function(x) { return(x[x['RCT_Assignment']]) }))
head(thedata, n = 3)
```

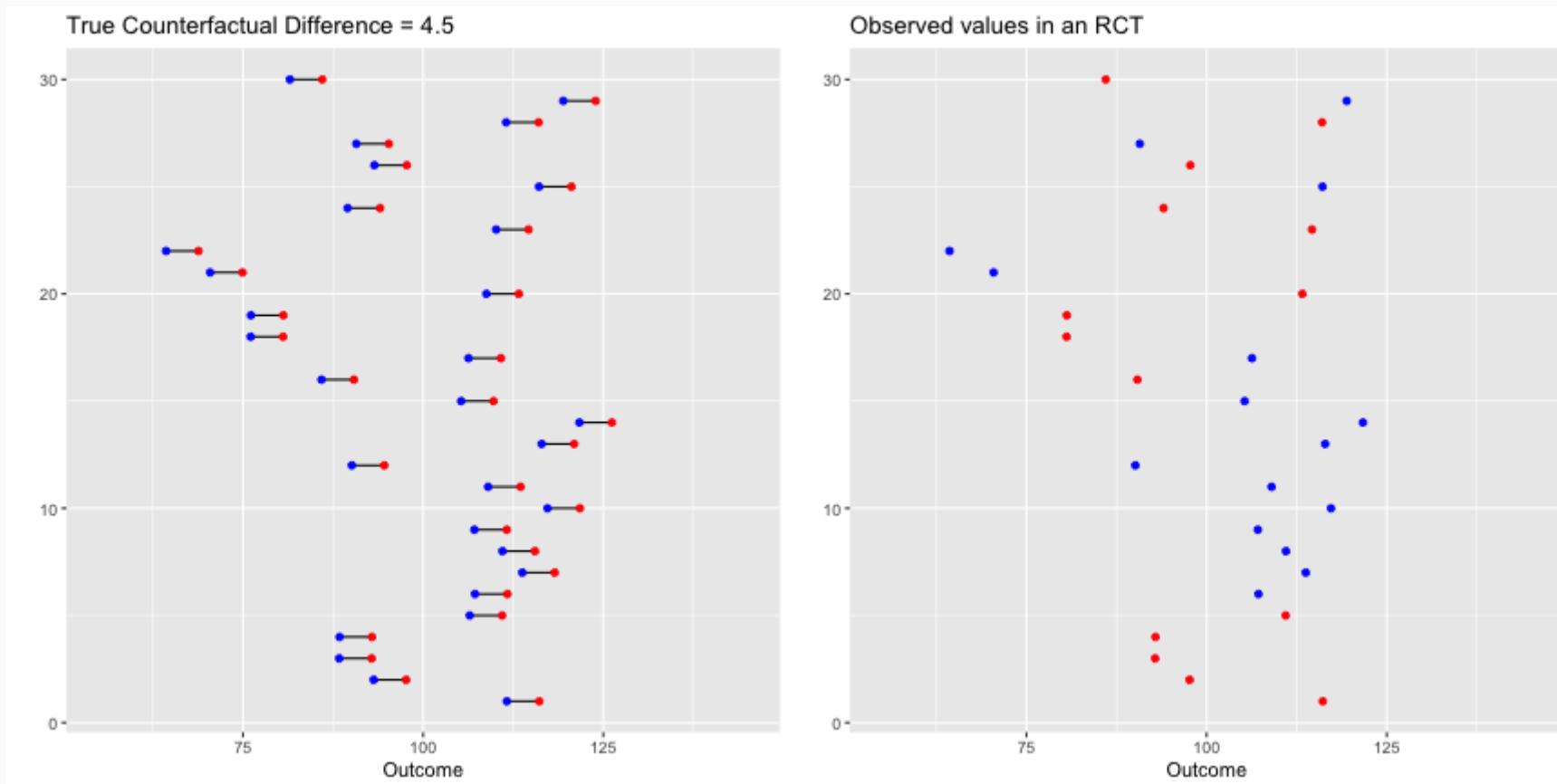
```
##   id   center   placebo treatment diff RCT_Assignment RCT_Value
## 1  1 113.86506 111.61506 116.11506  4.5      treatment 116.11506
## 2  2  95.38746  93.13746  97.63746  4.5      treatment  97.63746
## 3  3  90.60380  88.35380  92.85380  4.5      treatment  92.85380
```

```
tab.out <- describeBy(thedata$RCT_Value, group = thedata$RCT_Assignment, mat = TRUE, skew = FALSE)
```

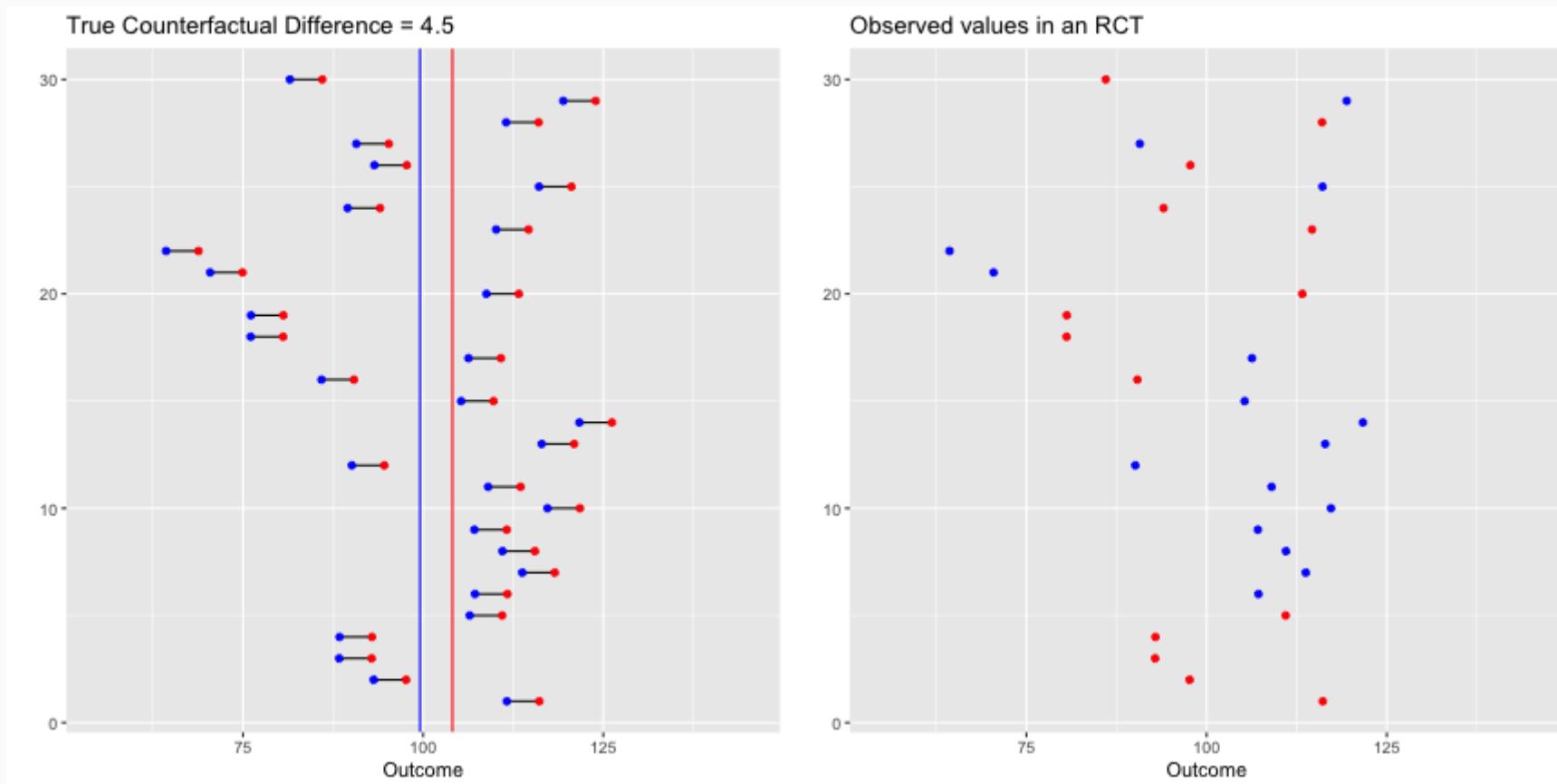
True Counterfactual



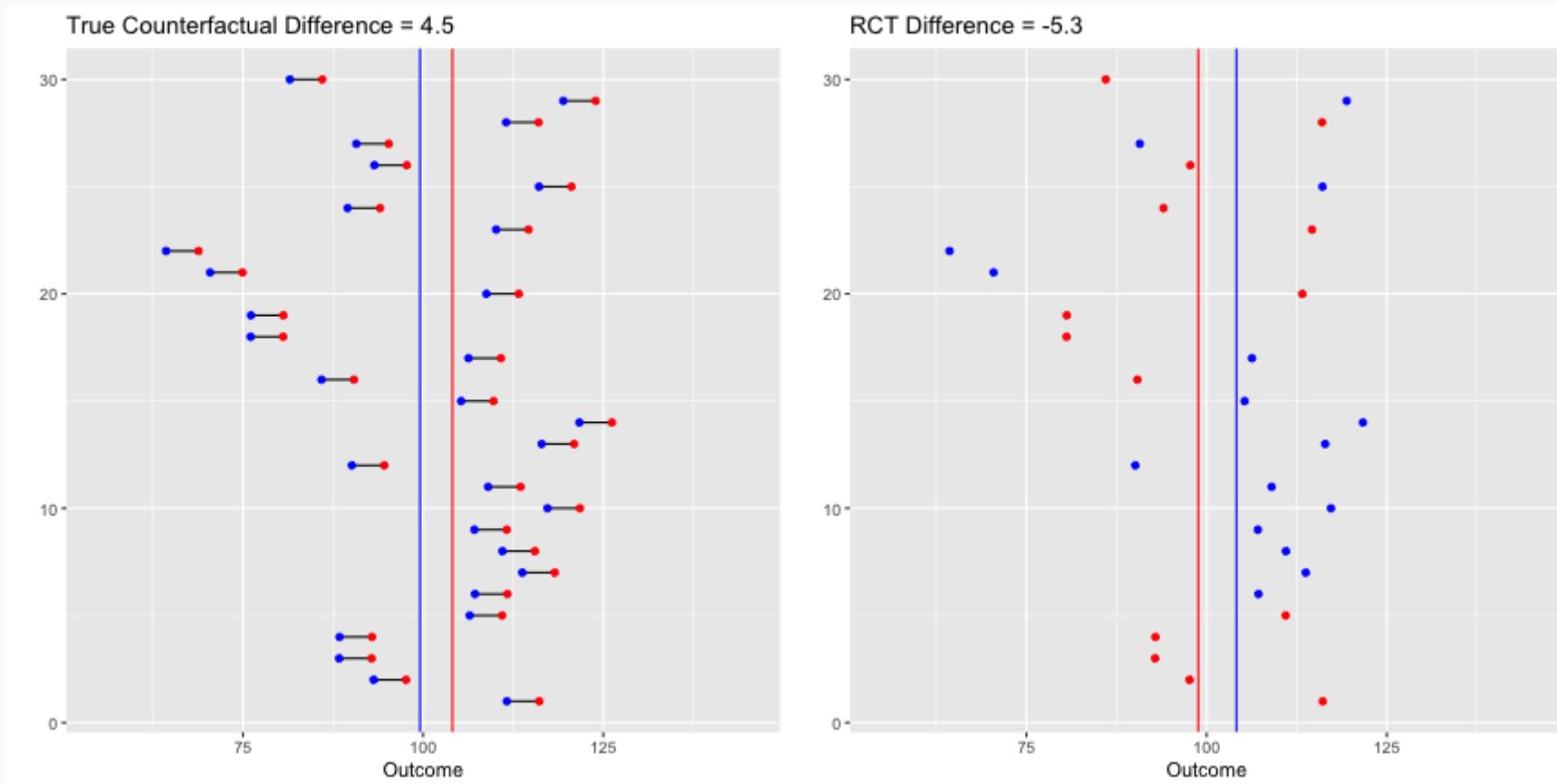
True Counterfactual (left) vs. One RCT (right)



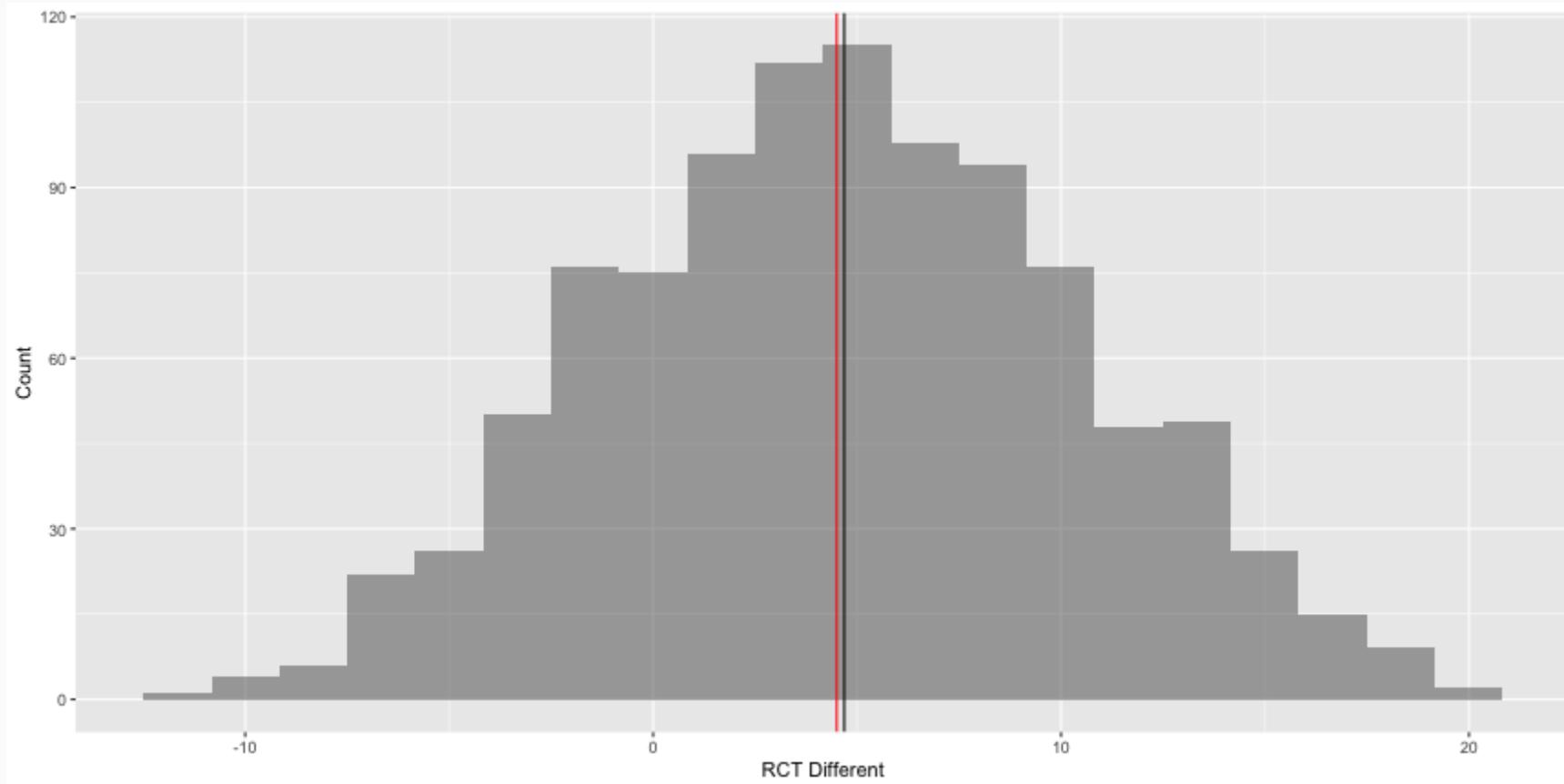
True Counterfactual (left) vs. One RCT (right)



True Counterfactual (left) vs. One RCT (right)



Distribution of Differences from 1,000 RCTs

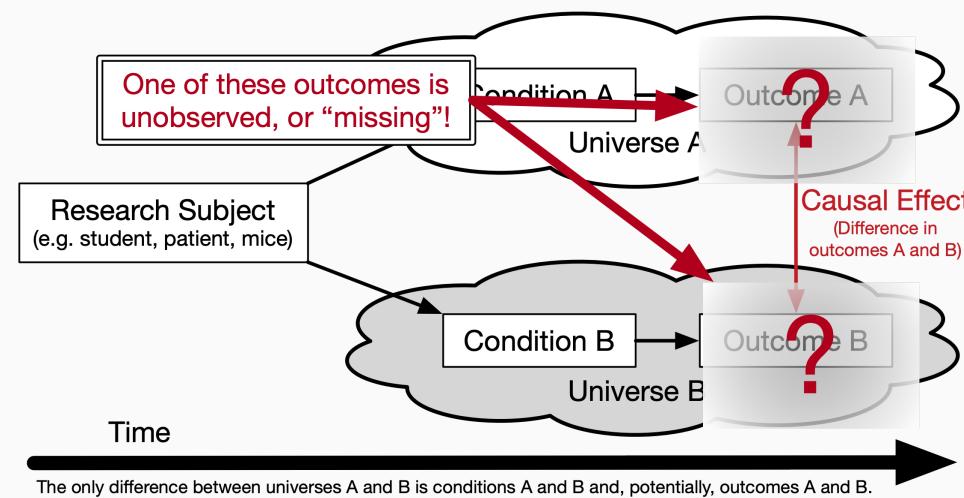


Rubin Causal Model

- The causal effect of a treatment is the difference in an individual's outcome under the situation they were given the treatment and not (referred to as a counterfactual).

$$\delta_i = Y_{i1} - Y_{i0}$$

- However, it is impossible to directly observe δ_i (referred to as *The Fundamental Problem of Causal Inference*, Holland 1986).
- Rubin frames this problem as a "missing data problem."



Propensity Score Analysis

The propensity score is the "conditional probability of assignment to a particular treatment given a vector of observed covariates" (Rosenbaum & Rubin, 1983, p. 41). The probability of being in the treatment:

$$\pi(X_i) \equiv Pr(T_i = 1 | X_i)$$

The balancing property under exogeneity:

$$T_i \perp\!\!\!\perp X_i | \pi(X_i)$$

We can then restate the ignorability assumption with the propensity score:

$$(Y_i(1), Y_i(0)) \perp\!\!\!\perp T_i | \pi(X_i)$$

Treatment Effects

The average treatment effect (ATE) is defined as:

$$E(r_1) - E(r_0)$$

where $E(\cdot)$ is the expectation in the population. For a set of covariates, X , and outcomes Y where 0 denotes control and 1 treatment, we define ATE as:

$$ATE = E(Y_1 - Y_0|X) = E(Y_1|X) - E(Y_0|X)$$

The Average treatment effect on the treated (ATT), is defined as:

$$ATT = E(Y_1 - Y_0|X, C = 1) = E(Y_1|X, C = 1) - E(Y_0|X, C = 1)$$



Propensity score methods

- **Matching** - Each treatment unit is paired with a comparison unit based upon the pre-treatment covariates.
- **Stratification** Treatment and comparison units are divided into strata (or subclasses) so that treated and comparison units are similar within each strata. Cochran (1968) observed that creating five subclassifications (stratum) removes at least 90\% of the bias in the estimated treatment effect.
- **Weighting** Each observation is weighted by the inverse of the probability of being in that group.

$$\frac{1}{n} \sum_{i=1}^n \left(\frac{T_i Y_i}{\pi(X_i)} - \frac{(1 - T_i) Y_i}{1 - \pi(X_i)} \right)$$

Steps for Implementing Matching Methods

Stuart and Rubin (2008) outline the following steps for matching, but the same approach can be used for stratification and weighting as well.

1. Choose the covariates to be used.
2. Define a distance measure (i.e. what constitutes similar).
3. Choose the matching algorithm.
4. Diagnose the matches (or strata) obtained (iterating through steps 2 and 3 as well).
5. Estimate the treatment effect using the matches (or strata) found in step 4.

Matching Methods

There are many choices and approaches to matching, including:

- Propensity score matching.
- Limited exact matching.
- Full matching.
- Nearest neighbor matching.
- Optimal/Genetic matching.
- Mahalanobis distance matching (for quantitative covariates only).
- Matching with and without replacement.
- One-to-one or one-to-many matching.

Which matching method should you use?

Whichever one gives the best balance!

See Rosenbaum (2012), *Testing one hypothesis twice in observational studies*.

Simulated Example

```
n <- 1000
treatment.effect <- 2
X <- mvtnorm::rmvnorm(n,
                      mean = c(0.5, 1),
                      sigma = matrix(c(2, 1, 1, 1), ncol = 2) )
dat <- tibble(
  x1 = X[, 1],
  x2 = X[, 2],
  treatment = as.numeric(- 0.5 + 0.25 * x1 + 0.75 * x2 + rnorm(n, 0, 1) > 0),
  outcome = treatment.effect * treatment + rnorm(n, 0, 1)
)
dat
```

```
## # A tibble: 1,000 x 4
##       x1      x2 treatment outcome
##   <dbl>  <dbl>     <dbl>    <dbl>
## 1  1.60   1.14      1     2.54
## 2 -0.619  0.163     0     0.544
## 3  1.56   1.82      0     0.207
## 4  2.33   2.27      1     1.53
## 5  1.92   2.44      1     2.07
## 6  1.28   0.880     1     1.43
## 7  2.89   2.98      1     1.36
```

Scatterplot of two covariates by treatment

```
ggplot(dat, aes(x = x1, y = x2, color = factor(treatment))) +  
  geom_point() + scale_color_manual('Treatment', values = cols)
```



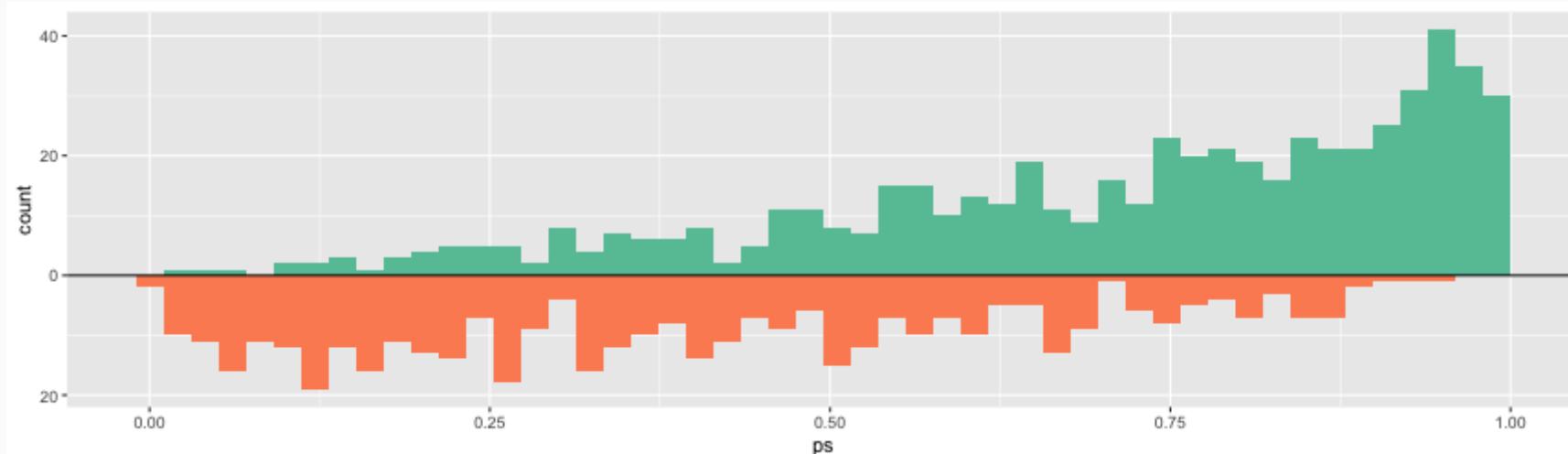
Estimate Propensity Scores

```
lr.out <- glm(treatment ~ x1 + x2, data = dat, family = binomial(link='logit'))
dat$ps <- fitted(lr.out) # Get the propensity scores
summary(lr.out)
```

```
## 
## Call:
## glm(formula = treatment ~ x1 + x2, family = binomial(link = "logit"),
##      data = dat)
## 
## Deviance Residuals:
##       Min        1Q     Median        3Q       Max
## -2.4355  -0.8136   0.3107   0.7797   2.7172
## 
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.07812   0.13066 -8.251 < 2e-16 ***
## x1          0.43117   0.08136  5.299 1.16e-07 ***
## x2          1.27902   0.12992  9.845 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## (Dispersion parameter for binomial family taken to be 1)
## 
```

Distribution of Propensity Scores

```
dat2 <- dat %>% tidyr::spread(treatment, ps, sep = '_p')
ggplot(dat) +
  geom_histogram(data = dat[dat$treatment == 1,], aes(x = ps, y = ..count..),
                 bins = 50, fill = cols[2]) +
  geom_histogram(data = dat[dat$treatment == 0,], aes(x = ps, y = -..count..),
                 bins = 50, fill = cols[1]) +
  geom_hline(yintercept = 0, lwd = 0.5) +
  scale_y_continuous(label = abs)
```

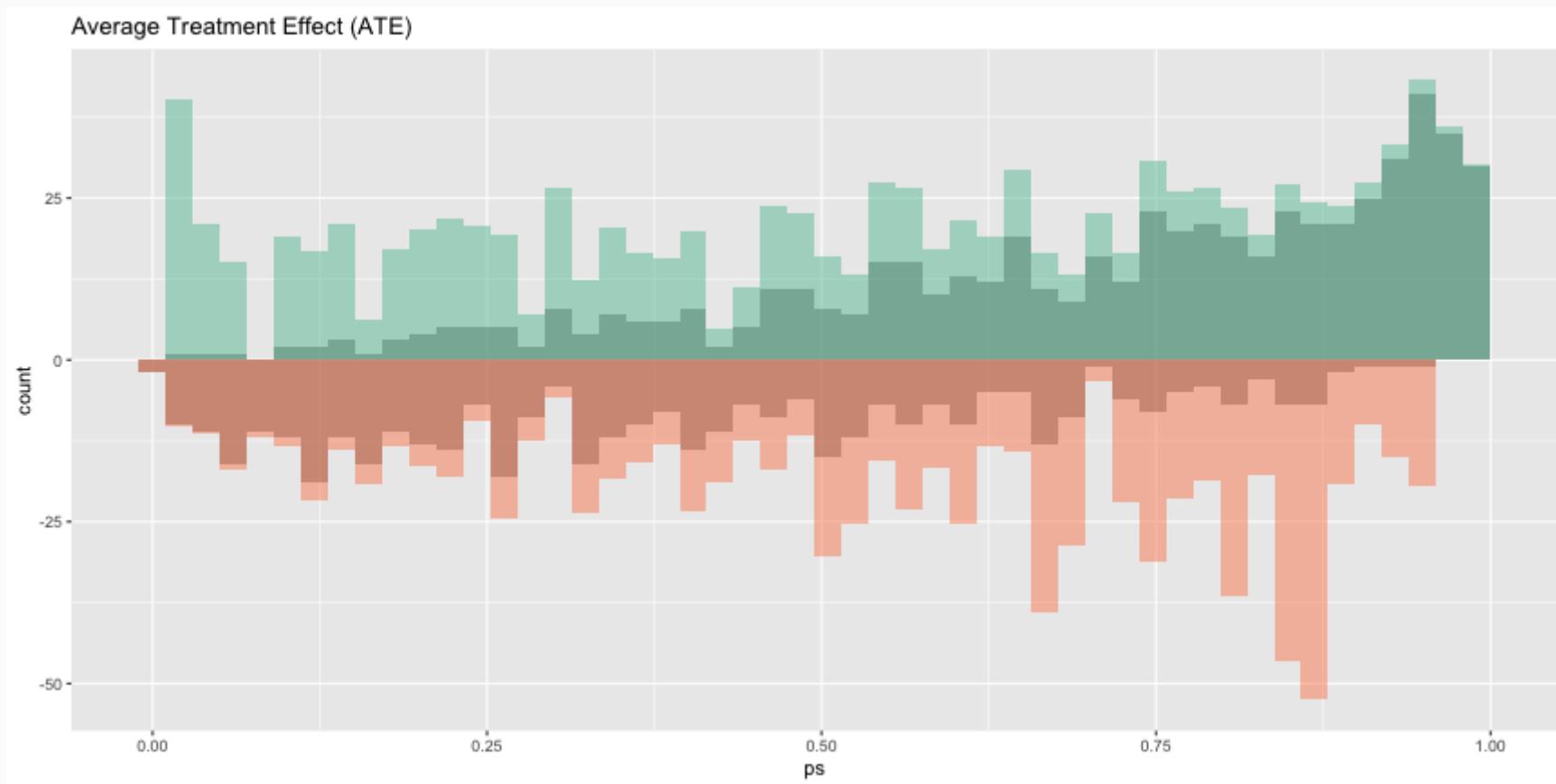


Propensity Score Weights

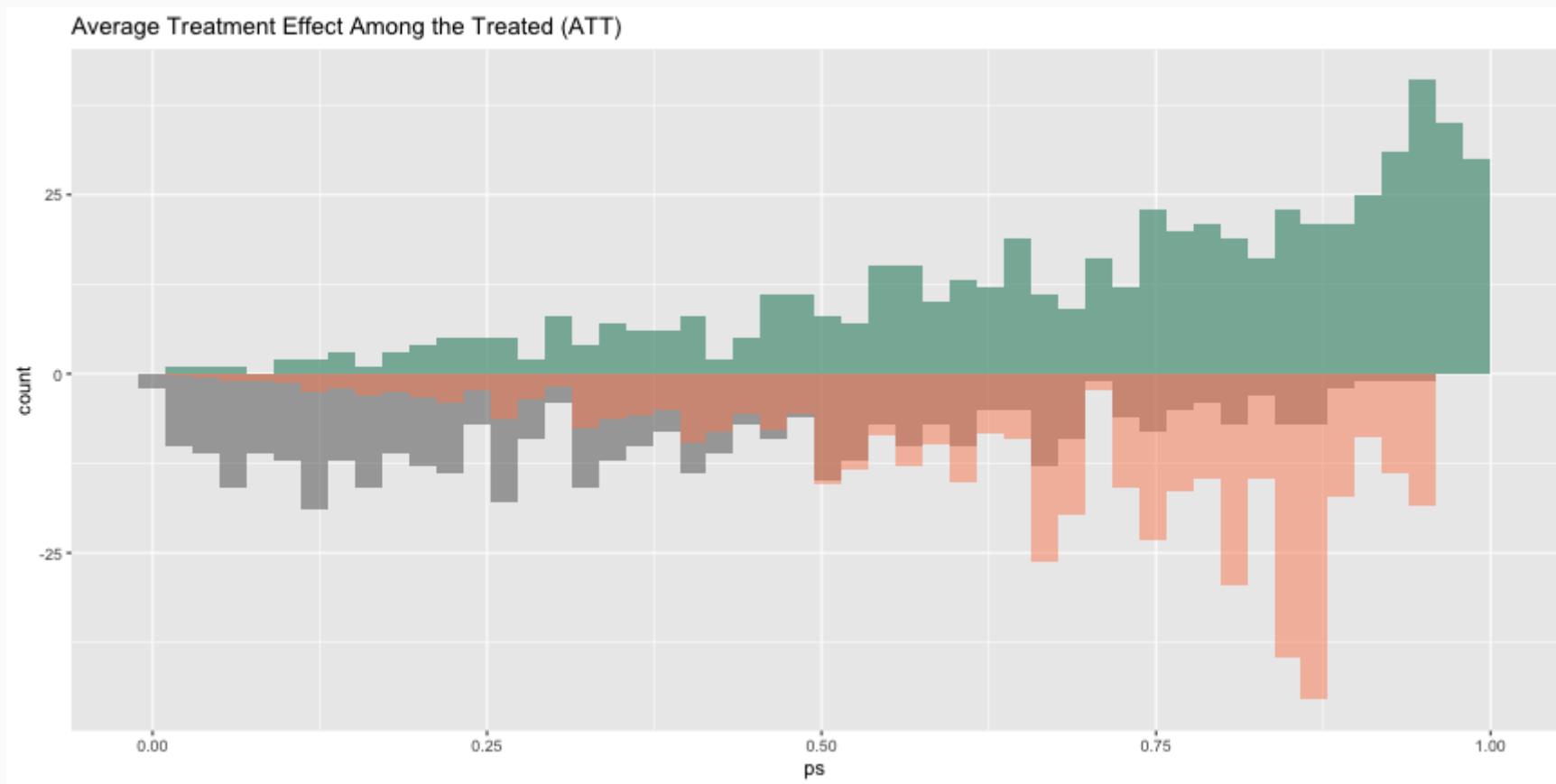
```
dat <- dat %>% mutate(  
    ate_weight = psa::calculate_ps_weights(treatment, ps, estimand = 'ATE'),  
    att_weight = psa::calculate_ps_weights(treatment, ps, estimand = 'ATT'),  
    atc_weight = psa::calculate_ps_weights(treatment, ps, estimand = 'ATC'),  
    atm_weight = psa::calculate_ps_weights(treatment, ps, estimand = 'ATM')  
)  
dat
```

```
## # A tibble: 1,000 x 9  
##       x1      x2 treatment outcome     ps ate_weight att_weight atc_weight atm_weight  
##   <dbl>  <dbl>     <dbl>    <dbl> <dbl>     <dbl>     <dbl>     <dbl>     <dbl>  
## 1  1.60  1.14        1     2.54  0.744     1.34      1       0.343     0.343  
## 2 -0.619  0.163       0     0.544  0.243     1.32      0.321     1       0.321  
## 3  1.56  1.82        0     0.207  0.873     7.86      6.86     1       1  
## 4  2.33  2.27        1     1.53   0.944     1.06      1       0.0593    0.0593  
## 5  1.92  2.44        1     2.07   0.946     1.06      1       0.0565    0.0565  
## 6  1.28  0.880       1     1.43   0.646     1.55      1       0.549     0.549  
## 7  2.89  2.98        1     1.36   0.982     1.02      1       0.0186   0.0186
```

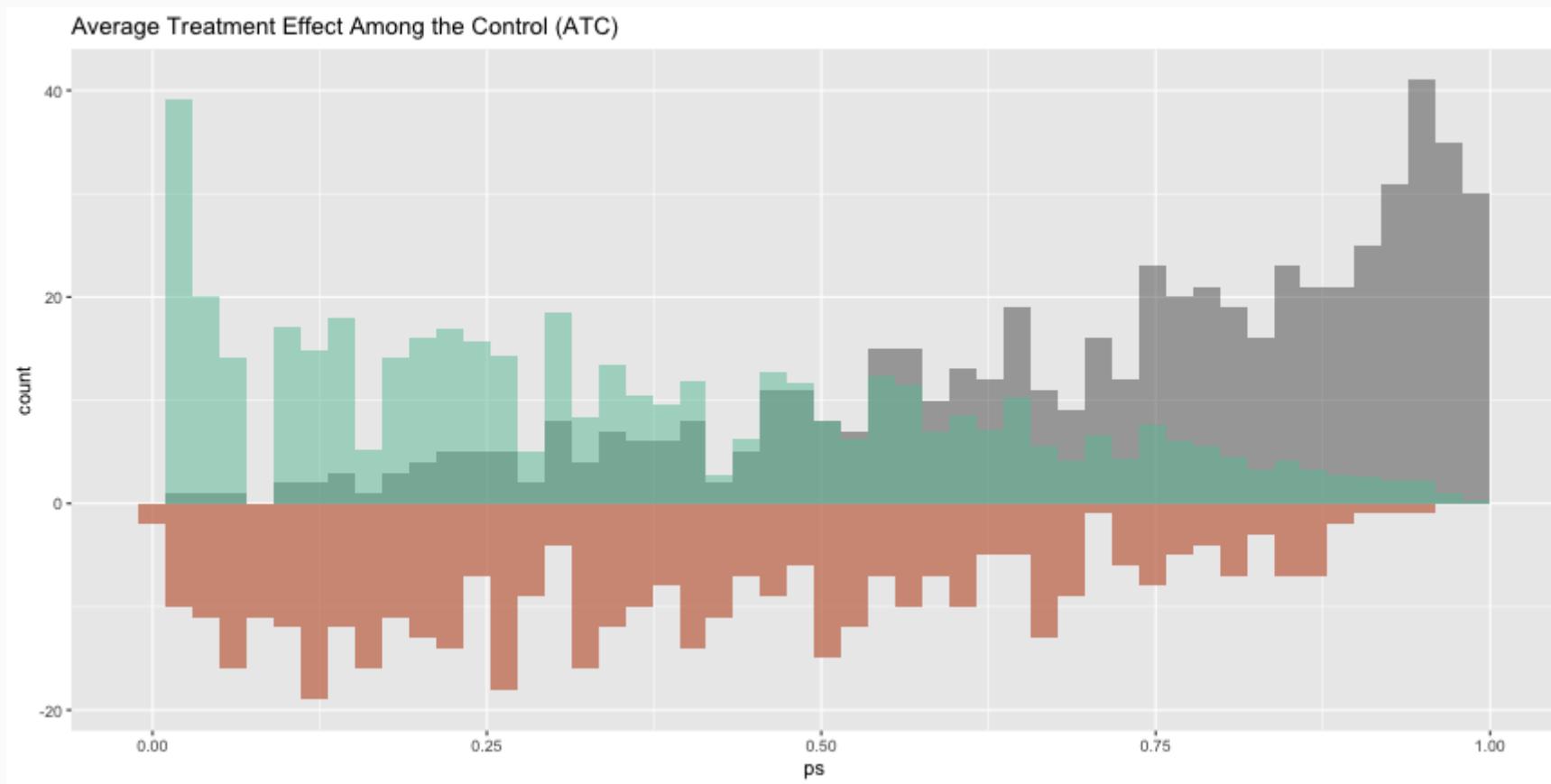
Average Treatment Effect (ATE)



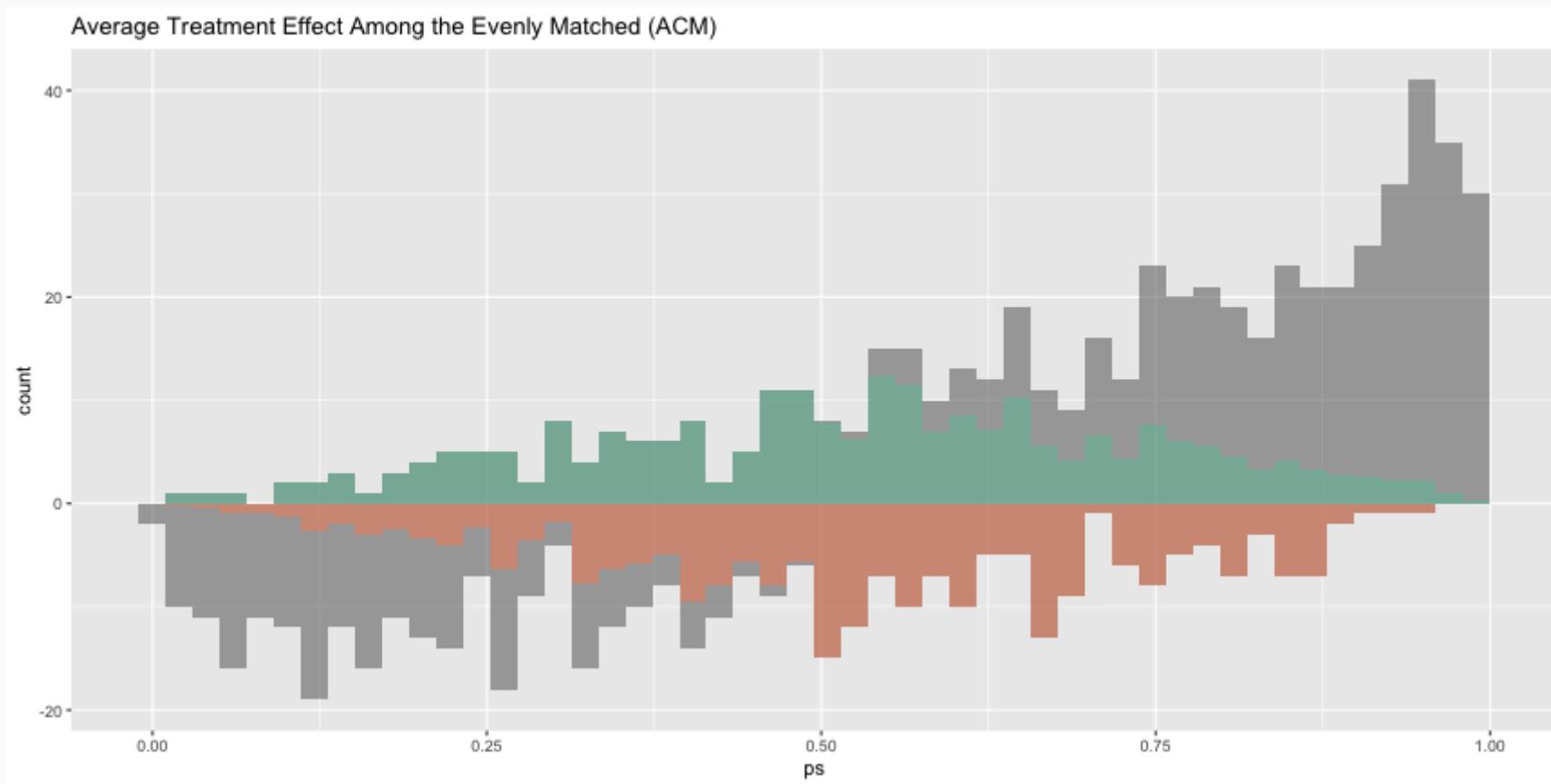
Average Treatment Effect Among the Treated (ATT)



Average Treatment Effect Among the Control (ATC)



Average Treatment Effect Among the Evenly Matched



National Supported Work

The National Supported Work (NSW) Demonstration was a federally and privately funded randomized experiment done in the 1970s to estimate the effects of a job training program for disadvantaged workers.

- Participants were randomly selected to participate in the training program.
- Both groups were followed up to determine the effect of the training on wages.
- Analysis of the mean differences (unbiased given randomization), was approximately \$800.

Lalonde (1986) used data from the Panel Survey of Income Dynamics (PSID) and the Current Population Survey (CPS) to investigate whether non-experimental methods would result in similar results to the randomized experiment. He found results ranging from \\$700 to \\$16,000.

National Supported Work (cont.)

Dehejia and Wahba (1999) later used propensity score matching to analyze the data. They found that,

- Comparison groups selected by Lalonde were very dissimilar to the treated group.
- By restricting the comparison group to those that were similar to the treated group, they could replicate the original NSW results.
- Using the CPS data, the range of treatment effect was between \$1,559 to \$1,681. The experimental results for the sample sample was approximately \$1,800.

The covariates available include: age, education level, high school degree, marital status, race, ethnicity, and earnings in 1974 and 1975.

Outcome of interest is earnings in 1978.

```
data(lalonde, package='Matching')
```



Estimating Propensity Scores

```
lalonde.formu <- treat~age + educ + black + hisp + married + nodegr + re74 + re75  
glm1 <- glm(lalonde.formu, family=binomial, data=lalonde)  
summary(glm1)
```

```
##  
## Call:  
## glm(formula = lalonde.formu, family = binomial, data = lalonde)  
##  
## Deviance Residuals:  
##      Min        1Q    Median        3Q       Max  
## -1.4358  -0.9904  -0.9071   1.2825   1.6946  
##  
## Coefficients:  
##              Estimate Std. Error z value Pr(>|z|)  
## (Intercept) 1.178e+00 1.056e+00  1.115  0.26474  
## age         4.698e-03 1.433e-02  0.328  0.74297  
## educ        -7.124e-02 7.173e-02 -0.993  0.32061  
## black        -2.247e-01 3.655e-01 -0.615  0.53874  
## hisp        -8.528e-01 5.066e-01 -1.683  0.09228 .  
## married     1.636e-01 2.769e-01  0.591  0.55463  
## nodegr      -9.035e-01 3.135e-01 -2.882  0.00395 **  
## re74        -3.161e-05 2.584e-05 -1.223  0.22122  
## re75        6.161e-05 4.358e-05  1.414  0.15744
```

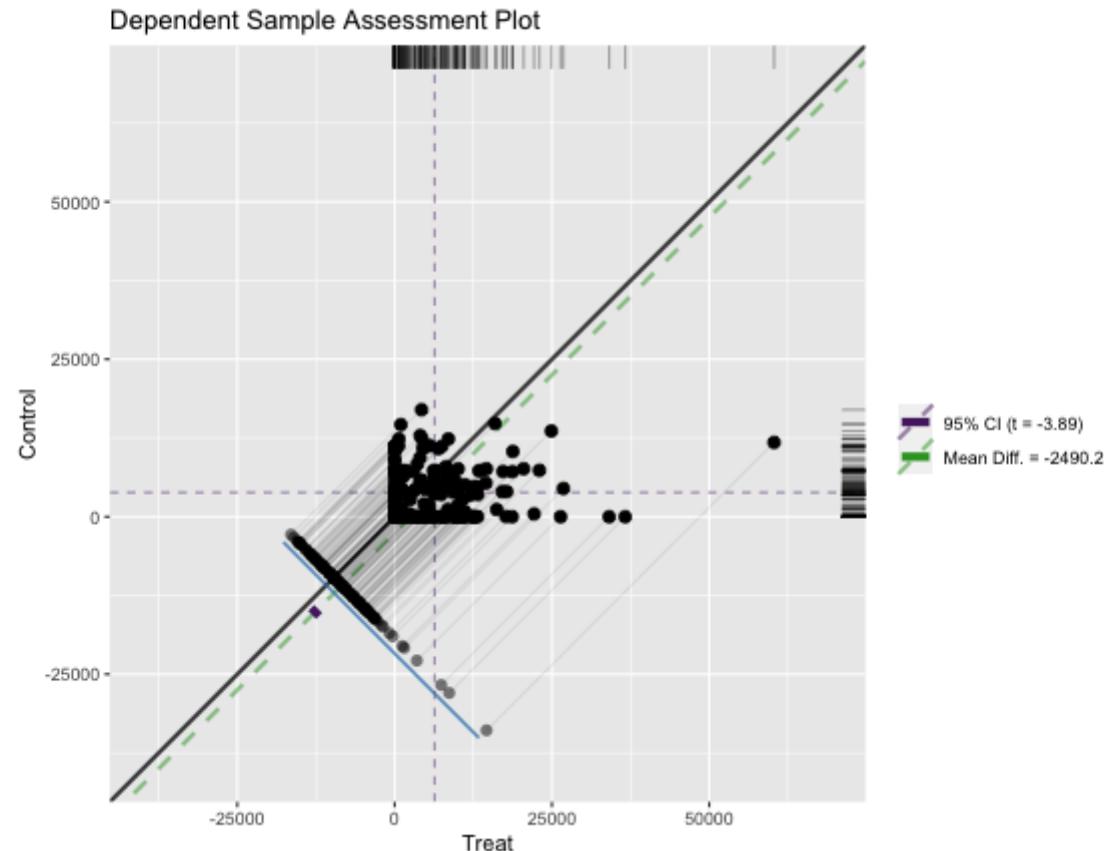
Estimating Propensity Scores

```
ps <- fitted(glm1) # Propensity scores
Y <- lalonde$re78 # Dependent variable, real earnings in 1978
Tr <- lalonde$treat # Treatment indicator
rr <- Match(Y=Y, Tr=Tr, X=ps, M=1, ties=FALSE)
summary(rr) # The default estimate is ATT here
```

```
## 
## Estimate... 2490.3
## SE......... 638.99
## T-stat..... 3.8972
## p.val..... 9.7305e-05
##
## Original number of observations..... 445
## Original number of treated obs..... 185
## Matched number of observations..... 185
## Matched number of observations (unweighted). 185
```

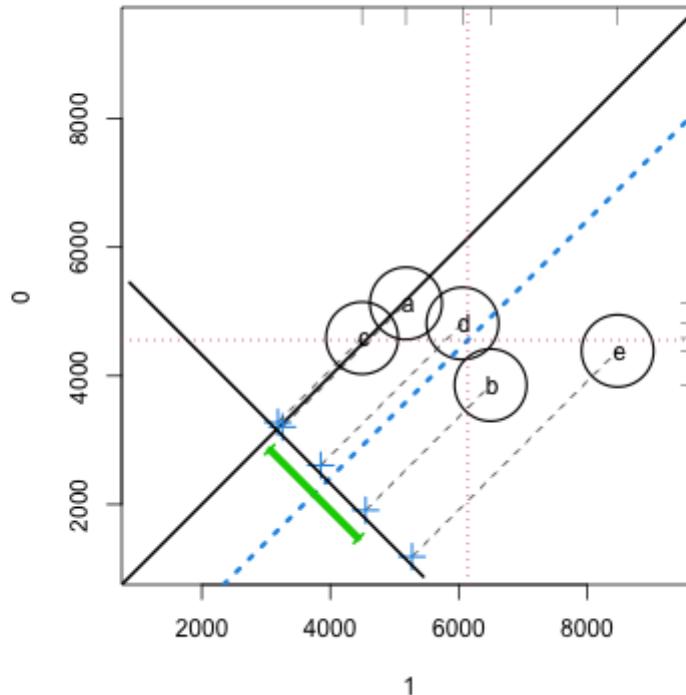
Visualizing Results

```
matches <- data.frame(Treat=lalonde[rr$index.treated,'re78'],  
                      Control=lalonde[rr$index.control,'re78'])  
granovagg.ds(matches[,c('Control','Treat')], xlab = 'Treat', ylab = 'Control')
```



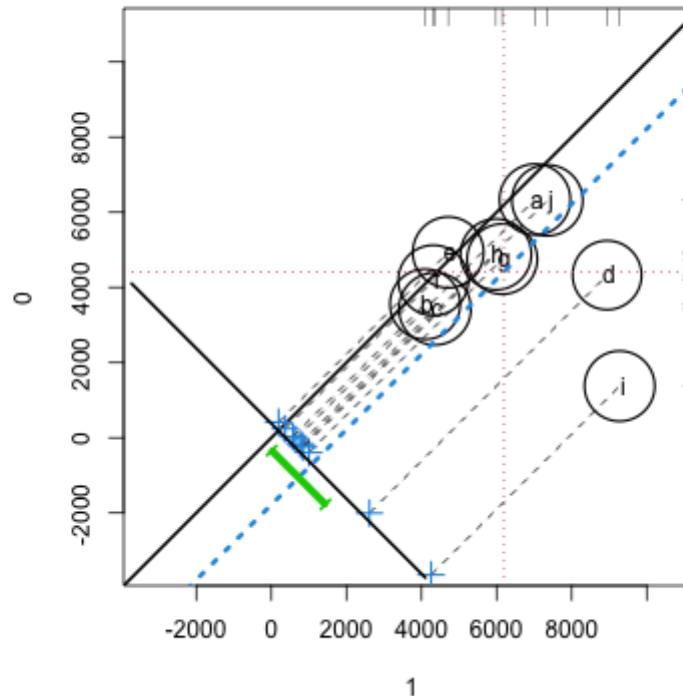
Stratification (5 Strata)

```
strata <- cut(ps, quantile(ps, seq(0, 1, 1/5)), include.lowest=TRUE, labels=letters[1:5])
circ.psa(lalonde$re78, lalonde$treat, strata, revc=TRUE)
```

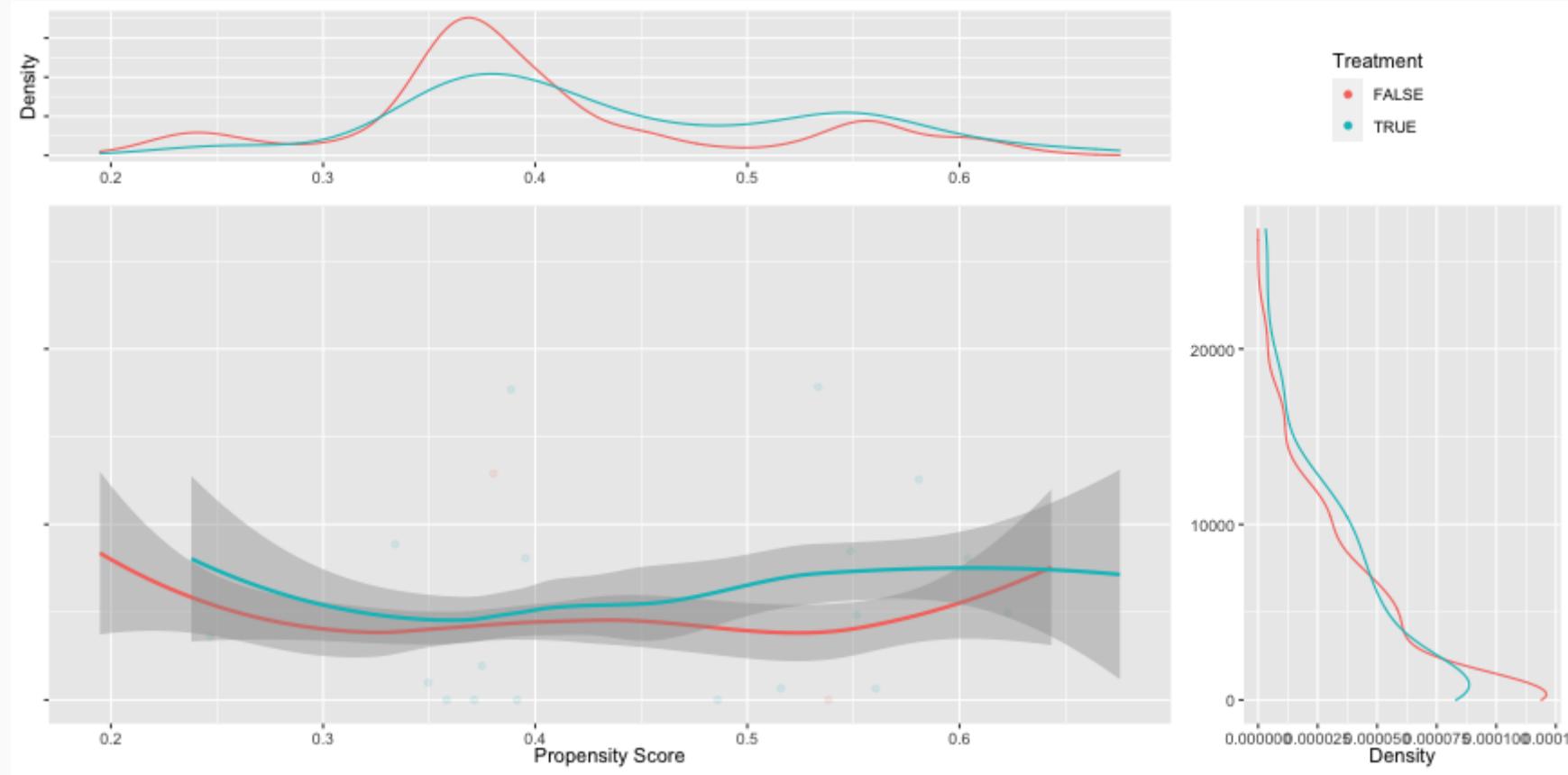


Stratification (10 Strata)

```
strata10 <- cut(ps, quantile(ps, seq(0, 1, 1/10)), include.lowest=TRUE, labels=letters[1:10])  
circ.psa(lalonde$re78, lalonde$treat, strata10, revc=TRUE)
```

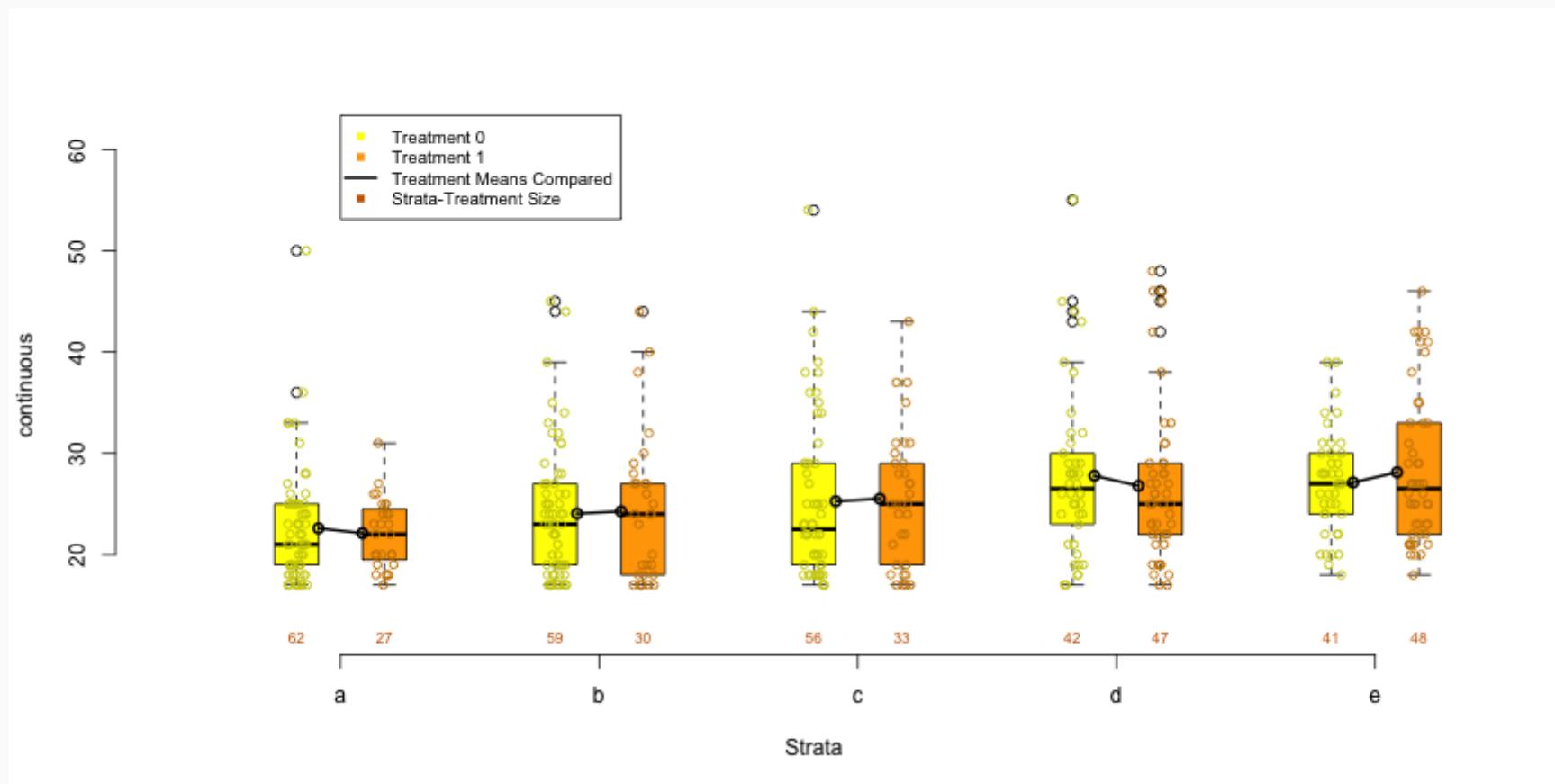


Loess Regression



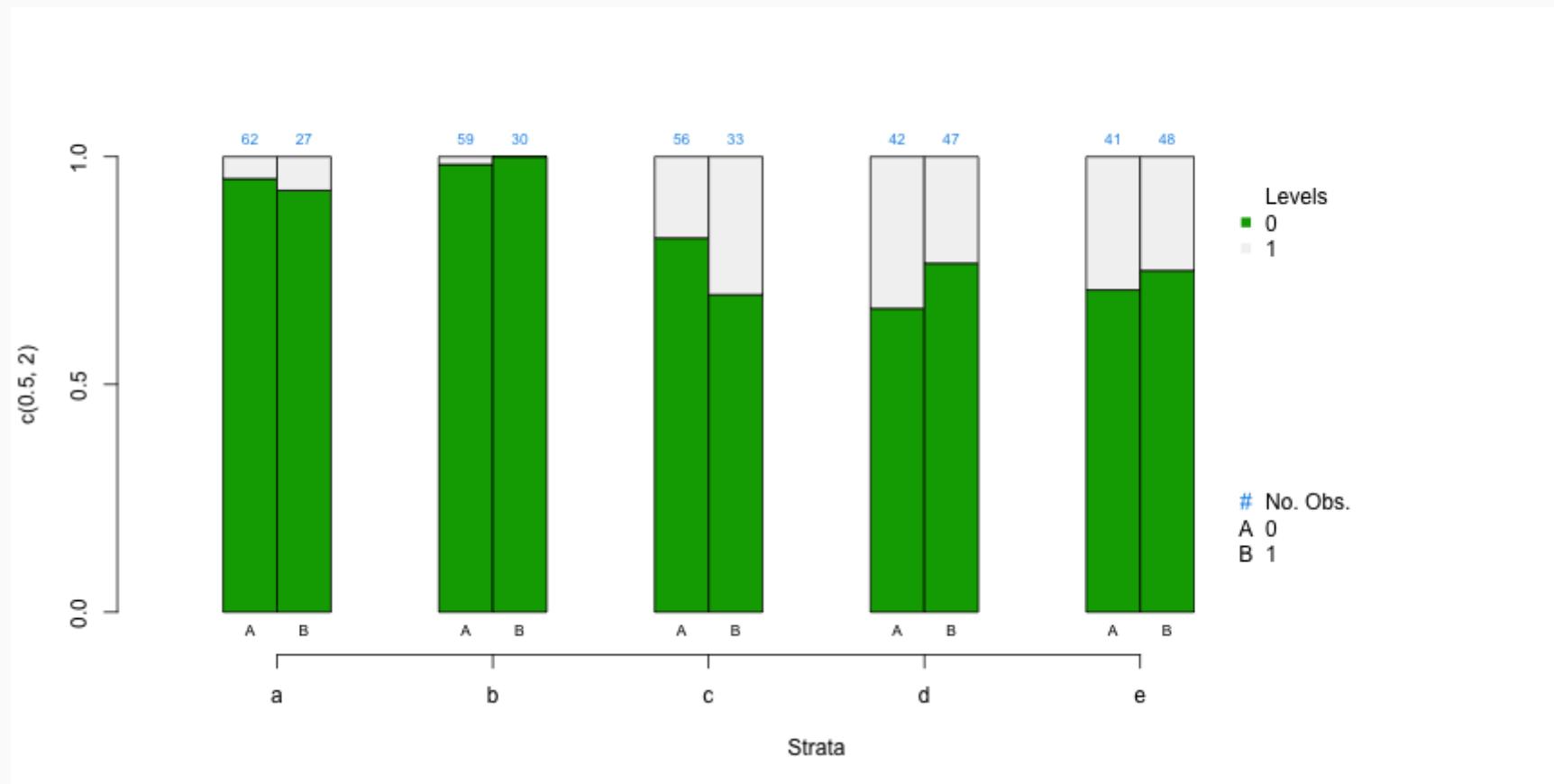
Checking Balance: Continuous Covariates

```
box.psa(lalonde$age, lalonde$treat, strata, xlab="Strata",  
balance=FALSE)
```

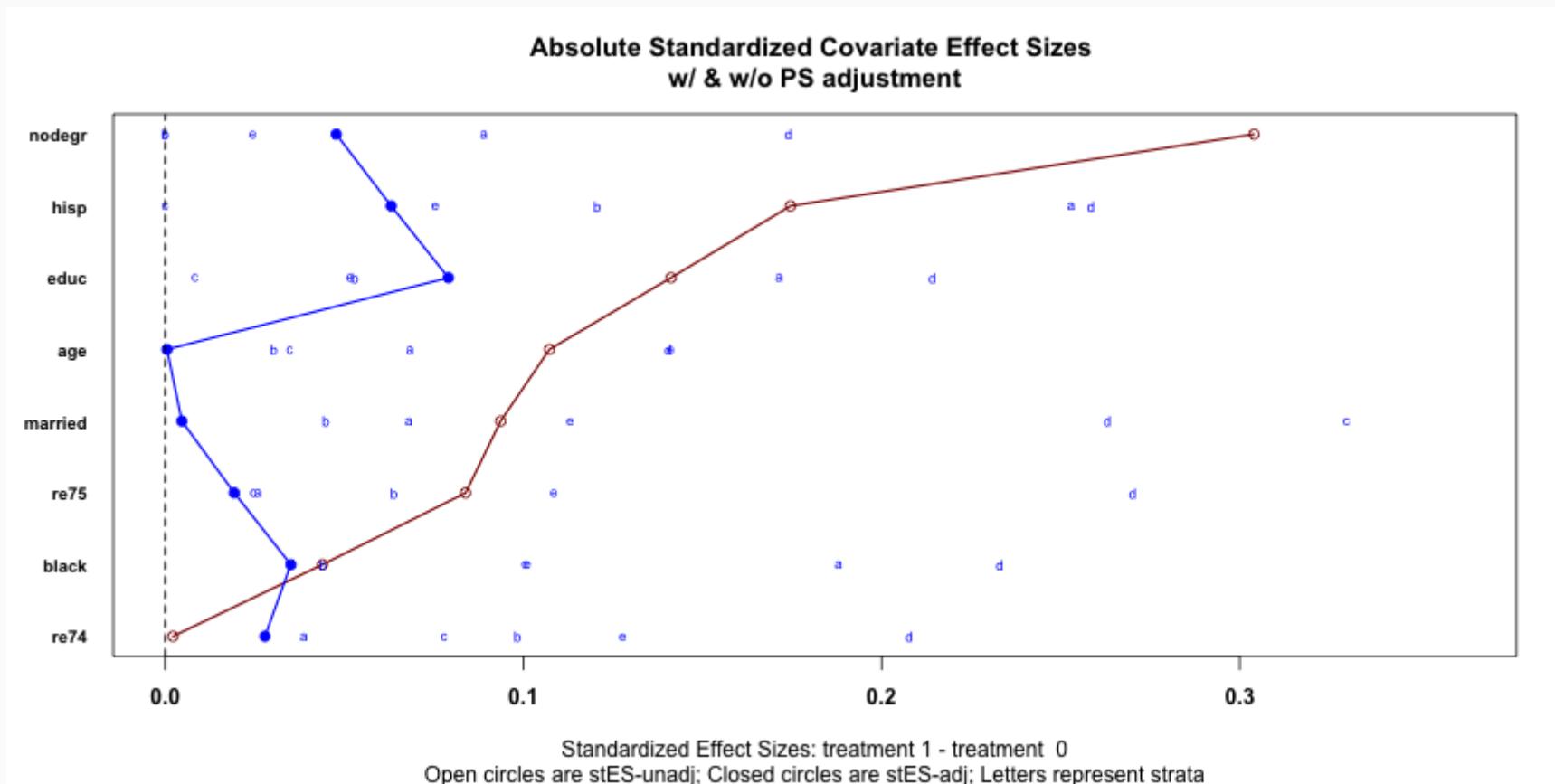


Checking Balance: Categorical Covariates

```
cat.psa(lalonde$married, lalonde$treat, strata, xlab='Strata',  
balance=FALSE)
```



Checking Balance: Covariate Balance Plot



Sensitivity Analysis

- An observational study is free of hidden bias if the propensity scores for each subject depend only on the observed covariates.
- That is, the p -value is valid *if* there are no unobserved confounders.
- However, there are very likely covariates that would better model treatment. These introduce hidden bias.
- Hidden bias exists if two subjects have the same covariates, but different propensity scores.

$X_a = X_b$ but $\pi_a \neq \pi_b$ for some a and b.

Sensitivity Analysis

Each person in the treatment is matched to exactly one person in the control. The odds of being in the treatment for persons a and b are:

$$O_a = \frac{\pi_a}{1-\pi_a} \text{ and } O_b = \frac{\pi_b}{1-\pi_b}$$

The ratio of these odds, Γ , measures the bias after matching.

$$\Gamma = \frac{O_a}{O_b} = \frac{\pi_a/(1 - \pi_a)}{\pi_b/(1 - \pi_b)}$$

This is the ratio of the odds the treated unit being in the treatment group to the matched control unit being in the treatment group.

Sensitivity Analysis

Sensitivity analysis tests whether the results hold for various ranges of Γ . That is, we test how large the differences in π (i.e. propensity scores) would have to be to change our basic inference. Let p_a and p_b be the probability of each unit of the matched pair being treated, conditional on exactly one being treated. For example:

- If $\Gamma = 1$, the treatment and control unit within each pair has the same value of treatment assignment ($p_a = 0.5$ and $p_b = 0.5$).
- If $\frac{1}{2} \leq \Gamma \leq 2$, no unit can be more than twice as likely as its match to get treated ($0.33 \leq p_a$, $p_b \leq 0.66$).
- If $\frac{1}{3} \leq \Gamma \leq 3$, no unit can be more than three times as likely as its match to get treated ($0.25 \leq p_a$, $p_b \leq 0.75$)

To get the bounds:

$$\frac{1}{\Gamma + 1} \leq p_a, p_b \leq \frac{\Gamma}{\Gamma + 1}$$



Wilcoxon Signed Rank Test

- Drop pairs where the matches have the same outcome.
- Calculate the difference in outcomes within each pair.
- Rank the pairs from smallest absolute difference to largest absolute difference (i.e. the smallest = 1).
- Take the sum of the ranks where the treated unit had the higher outcome.

$$W = \left| \sum_{1}^{N_r} sgn(x_{T,i} - x_{C,i}) \cdot R_i \right|$$

Where N is the number of ranked pairs; R_i is the rank for pair r ; $x_{T,i}$ and $x_{C,i}$ are the outcomes for the i^{th} treated and control pair, respectively.

Sensitivity Analysis

The process for sensitivity analysis:

- Select a series of values for Γ . For social science research, values between 1 and 2 is an appropriate start.
- For each Γ , estimate the p -values to see how the p -values increase for larger values of Γ .
- For binary outcomes, use McNemar's test, for all others use Wilcoxon sign rank test and the Hodges-Lehmann point estimate. See Keele (2010) for more information.

Sensitivity Analysis

Children of parents who had worked in a factory where lead was used in making batteries were matched by age, exposure to traffic, and neighborhood with children whose parents did not work in lead-related industries. Whole blood was assessed for lead content yielding measurements in mg/dl

```
require(rbounds)
trt <- c(38, 23, 41, 18, 37, 36, 23, 62, 31, 34, 24, 14, 21, 17,
       16, 20, 15, 10, 45, 39, 22, 35, 49, 48, 44, 35, 43, 39,
       34, 13, 73, 25, 27)
ctrl <- c(16, 18, 18, 24, 19, 11, 10, 15, 16, 18, 18, 13, 19,
        10, 16, 16, 24, 13, 9, 14, 21, 19, 7, 18, 19, 12,
        11, 22, 25, 16, 13, 11, 13)
```

Sensitivity Analysis

```
psens(trt, ctrl)
```

```
##  
## Rosenbaum Sensitivity Test for Wilcoxon Signed Rank P-Value  
##  
## Unconfounded estimate .... 0  
##  
## Gamma Lower bound Upper bound  
## 1 0 0.0000  
## 2 0 0.0018  
## 3 0 0.0131  
## 4 0 0.0367  
## 5 0 0.0694  
## 6 0 0.1073  
##  
## Note: Gamma is Odds of Differential Assignment To  
## Treatment Due to Unobserved Factors  
##
```

Sensitivity Analysis

```
hlsens(trt, ctrl)
```

```
##  
## Rosenbaum Sensitivity Test for Hodges-Lehmann Point Estimate  
##  
## Unconfounded estimate .... 15.5  
##  
## Gamma Lower bound Upper bound  
## 1 15.5 15.5  
## 2 10.5 19.6  
## 3 8.0 23.1  
## 4 6.5 25.1  
## 5 5.0 26.6  
## 6 4.0 28.1  
##  
## Note: Gamma is Odds of Differential Assignment To  
## Treatment Due to Unobserved Factors  
##
```

Matching of Non-Binary Treatments

- The **TriMatch** package provides functions for finding matched triplets.
- Estimates propensity scores for three separate logistic regression models (one for each pair of groups, that is, treat1-to-control, treat2-to-control, and treat1-to-treat2).
- Finds matched triplets that minimize the total distance (i.e. sum of the standardized distance between propensity scores within the three models). within a caliper.
- Provides multiple methods for determining which matched triplets are retained:
- Optimal which attempts to retain all treatment units.
- Full which retains all matched triplets within the specified caliper (.25 by default as suggested by Rosenbaum).
- Analog of the one-to-many for matched triplets. Specify how many times each treat1 and treat2 unit can be matched.
- Unique which allows each unit to be matched once, and only once.
- Functions for conducting repeated measures ANOVA and Freidman Ranksum Tests are provided.



Example: Tutoring

Students can opt to utilize tutoring services to supplement math courses. Of those who used tutoring services, approximately 58% of students used the tutoring service once, whereas the remaining 42% used it more than once. Outcome of interest is course grade.

- **Military** Active military status.
- **Income** Income level.
- **Employment** Employment level.
- **NativeEnglish** Is English their native language
- **EdLevelMother** Education level of their mother.
- **EdLevelFather** Education level of their father.
- **Ethnicity** American Indian or Alaska Native, Asian, Black or African American, Hispanic, Native Hawaiian or Other Pacific Islander, Two or more races, Unknown, White
- **Gender** Male, Female
- **Age** Age at course start.
- **GPA** Student GPA at the beginning of the course.

New Student Outreach: Covariates

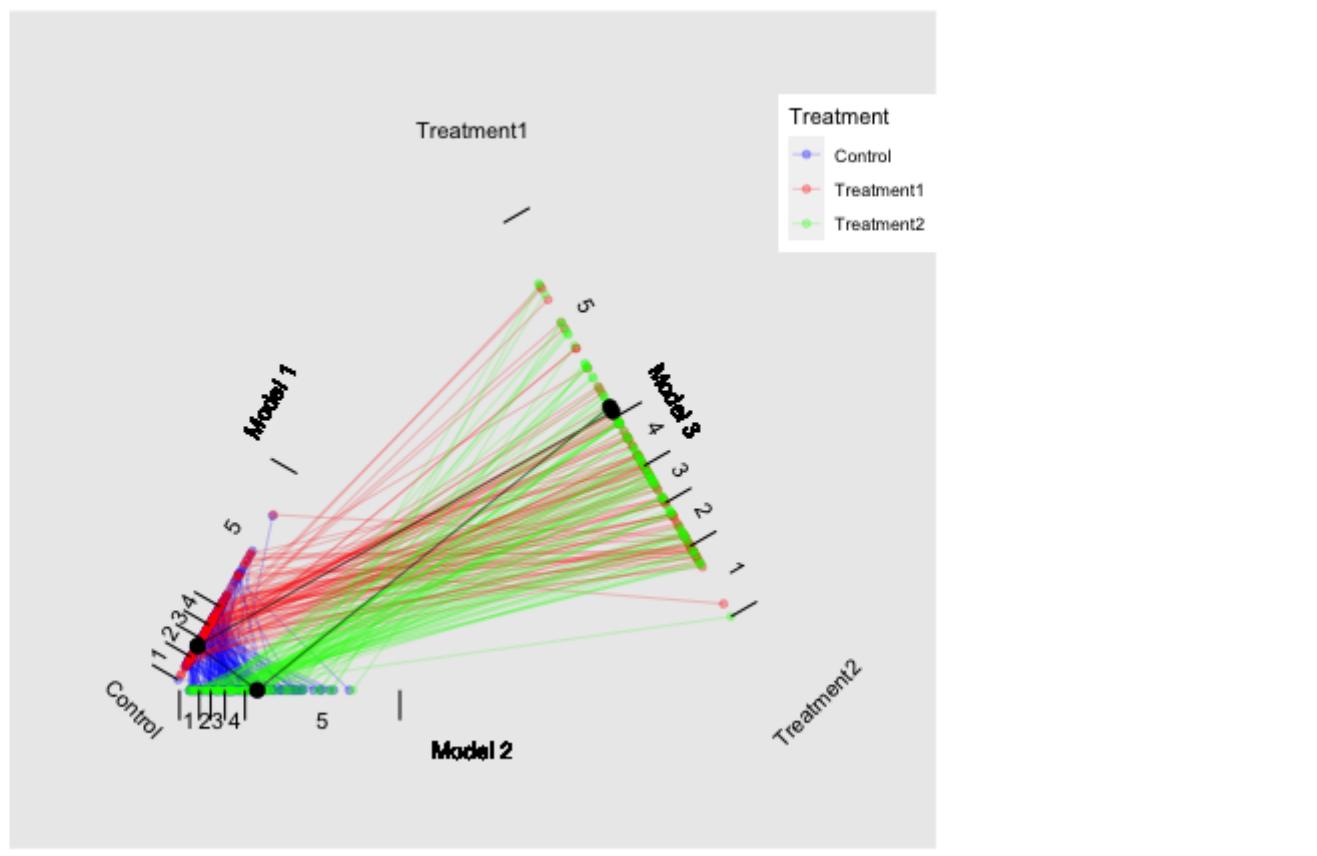
Newly enrolled students received outreach contacts until they registered for a course or six months have passed, whichever came first. Outreach was conducted by two academic advisors and a comparison group was drawn from students who enrolled prior to the start of the outreach program. Outcome of interest is number of credits attempted within the first seven months of enrollment.

PSA for Non-Binary Treatments

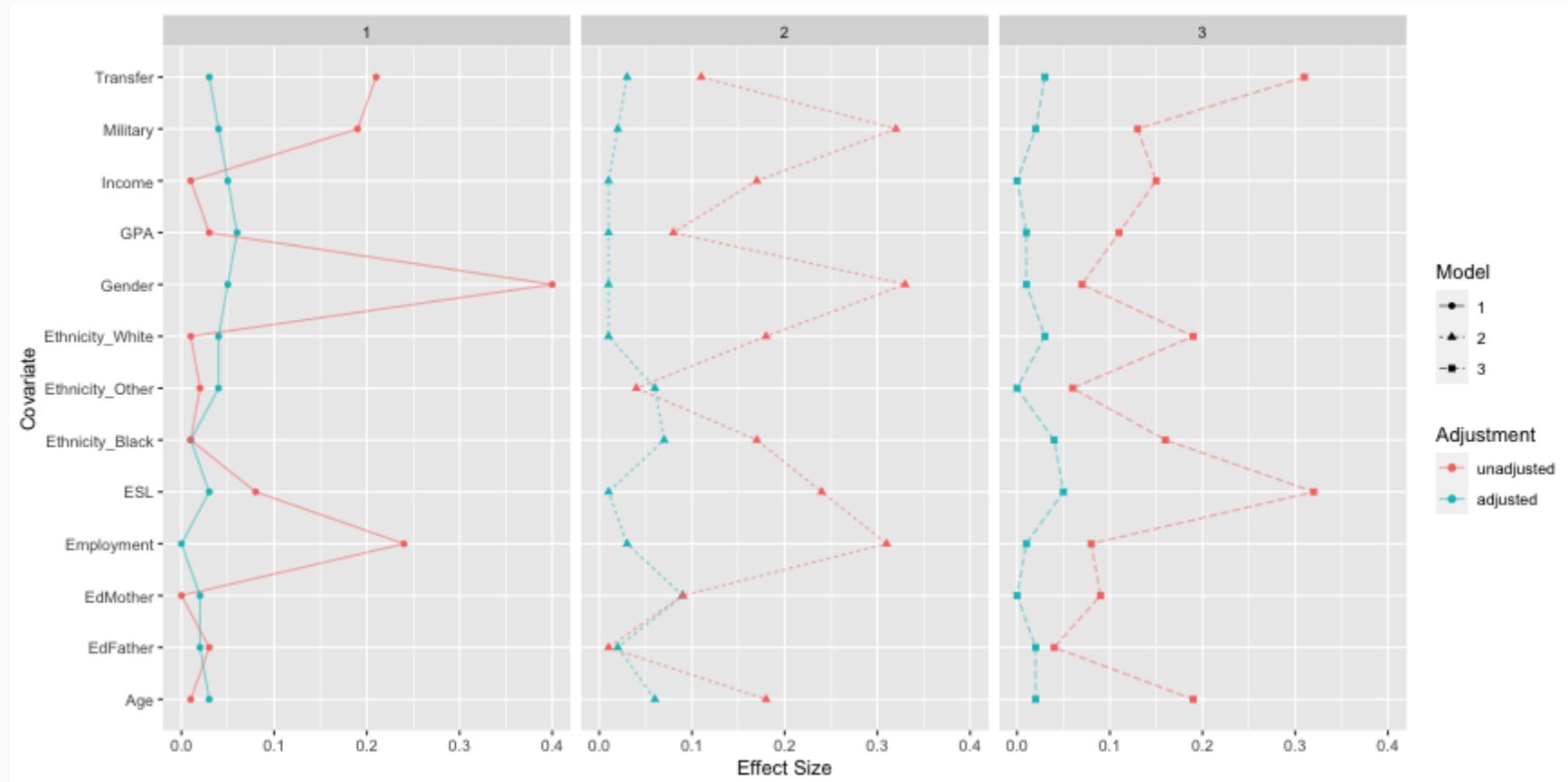
The **TriMatch** algorithm works as follows:

1. Estimate three separate propensity score models for each pair of groups (i.e. Control-to-Treat1, Control-to-Treat2, Treat1-to-Treat2).
2. Determine the matching order. The default is to start with the largest of two treatments, then the other treatment, followed by the control.
3. For each unit in group 1, find all units from group 2 within a certain threshold (i.e. difference between PSs is within a specified caliper).
4. For each unit in group 2, find all units from group 3 within a certain threshold.
5. Calculate the distance (difference) between each unit 3 found and the original unit 1. Eliminate candidates that exceed the caliper.
6. Calculate a total distance (sum of the three distances) and retain the smallest unique M group 1 units (by default $M=2$)

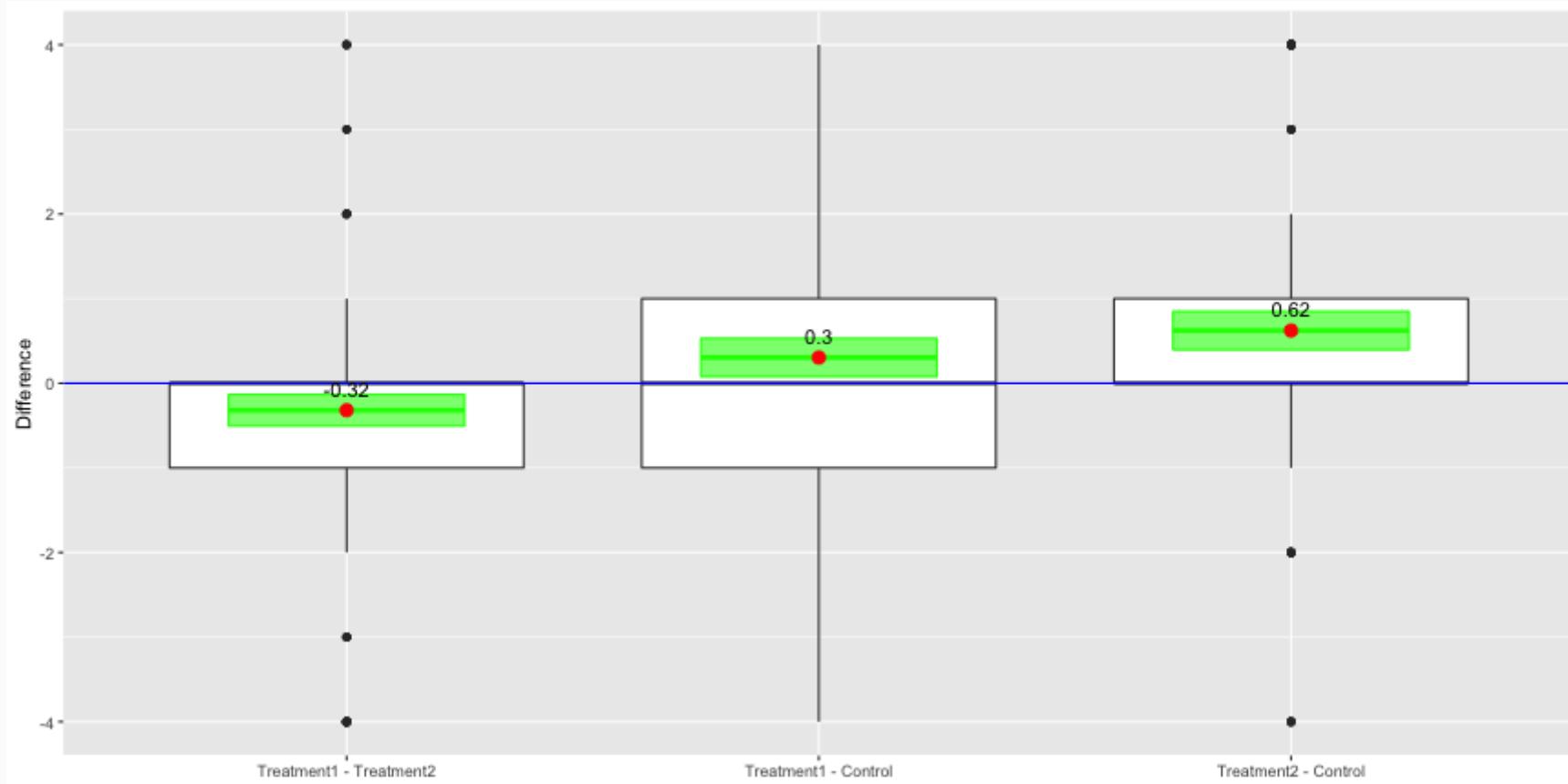
Matching Triplets



Checking Balance



Results



Multilevel PSA

The use of PSA for clustered, or multilevel data, has been limited (Thoemmes \& Felix, 2011). Bryer and Pruzek (2012, 2013) have introduced an approach to analyzing multilevel or clustered data using stratification methods and implemented in the `multilevelPSA` R package.

- Exact and partially exact matching methods implicitly adjust for clustering. That is, the covariates chosen to exactly match are, in essence, clustering variables.
- Exact matching only applies to phase I of PSA. How are the clusters related to outcome of interest.

The `multilevelPSA` uses stratification methods (e.g. quintiles, classification trees) by:

- Estimate separate propensity scores for each cluster.
- Identify strata within each cluster (e.g. leaves of classification trees, quintiles).
- Estimate ATE (or ATT) within each cluster.
- Aggregate estimated ATE to provide an overall ATE estimate.
- Several functions to summarize and visualize results and check balance.

The Programme of International Student Assessment

- International assessment conducted by the Organization for Economic Co-operation and Development (OECD).
- Assesses students towards the end of secondary school (approximately 15-year-old children) in math, reading, and science.
- Collects a robust set of background information from students, parents, teachers, and schools.
- Assess both private and public school students in many countries.
- We will use PISA to estimate the effects of private school attendance on PISA outcomes.

Phase I of Multilevel PSA

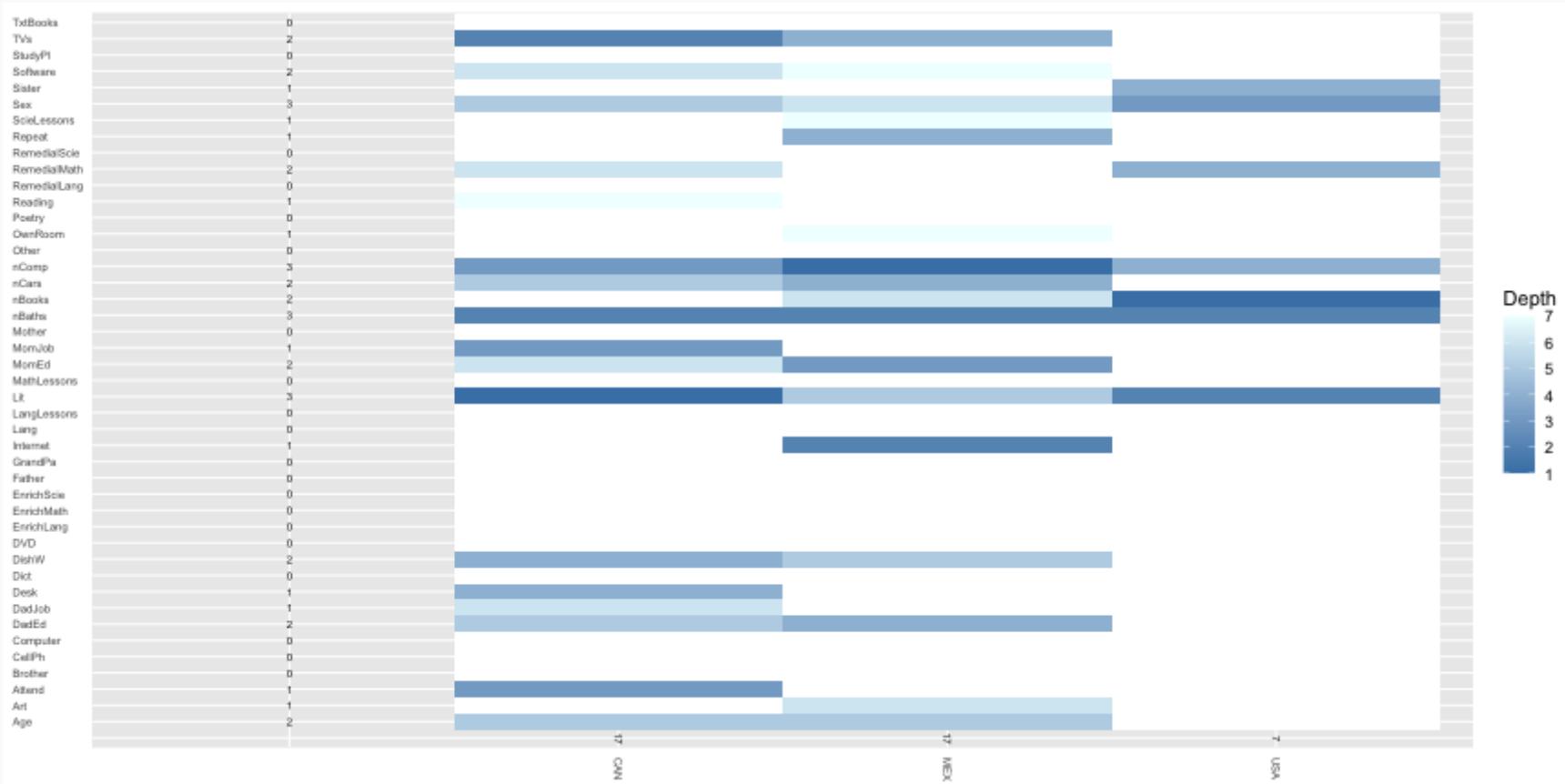
The `multilevelPSA` provides two functions, `mlpsa.ctree` and `mlpsa.logistic`, that will estimate propensity scores using classification trees and logistic regression, respectively. Since logistic regression requires a complete dataset (i.e. no missing values), we will use classification trees in this example.

```
data(pisana)
data(pisa.colnames)
data(pisa.psa.cols)
student = pisana
mlctree = mlpsa.ctree(student[,c('CNT', 'PUBPRIV', pisa.psa.cols)],
                      formula=PUBPRIV ~ ., level2='CNT')
student.party = getStrata(mlctree, student, level2='CNT')
student.party$mathscore = apply(
  student.party[,paste0('PV', 1:5, 'MATH')], 1, sum) / 5
```

To assess what covariates were used in each tree model, as well as the relative importance, we can create a heat map of covariate usage by level.

Covariate Heat Map

```
tree.plot(mlctree, level2Col=student$CNT, colLabels=pisa.colnames[,c('Variable','ShortDesc')])
```



Phase II of Multilevel PSA

The `mlpsa` function will compare the outcome of interest.

```
results.psa.math = mlpsa(response=student.party$mathscore,  
                           treatment=student.party$PUBPRIV, strata=student.party$strata,  
                           level2=student.party$CNT, minN=5)
```

```
results.psa.math$overall.wtd
```

```
## [1] -28.02406
```

```
results.psa.math$overall.ci
```

```
## [1] -31.30261 -24.74552
```

```
results.psa.math$level2.summary[,c('level2','Private','Private.n',  
'Public','Public.n','diffwtd','ci.min','ci.max')]
```

```
##   level2 Private Private.n   Public Public.n    diffwtd    ci.min    ci.max
```

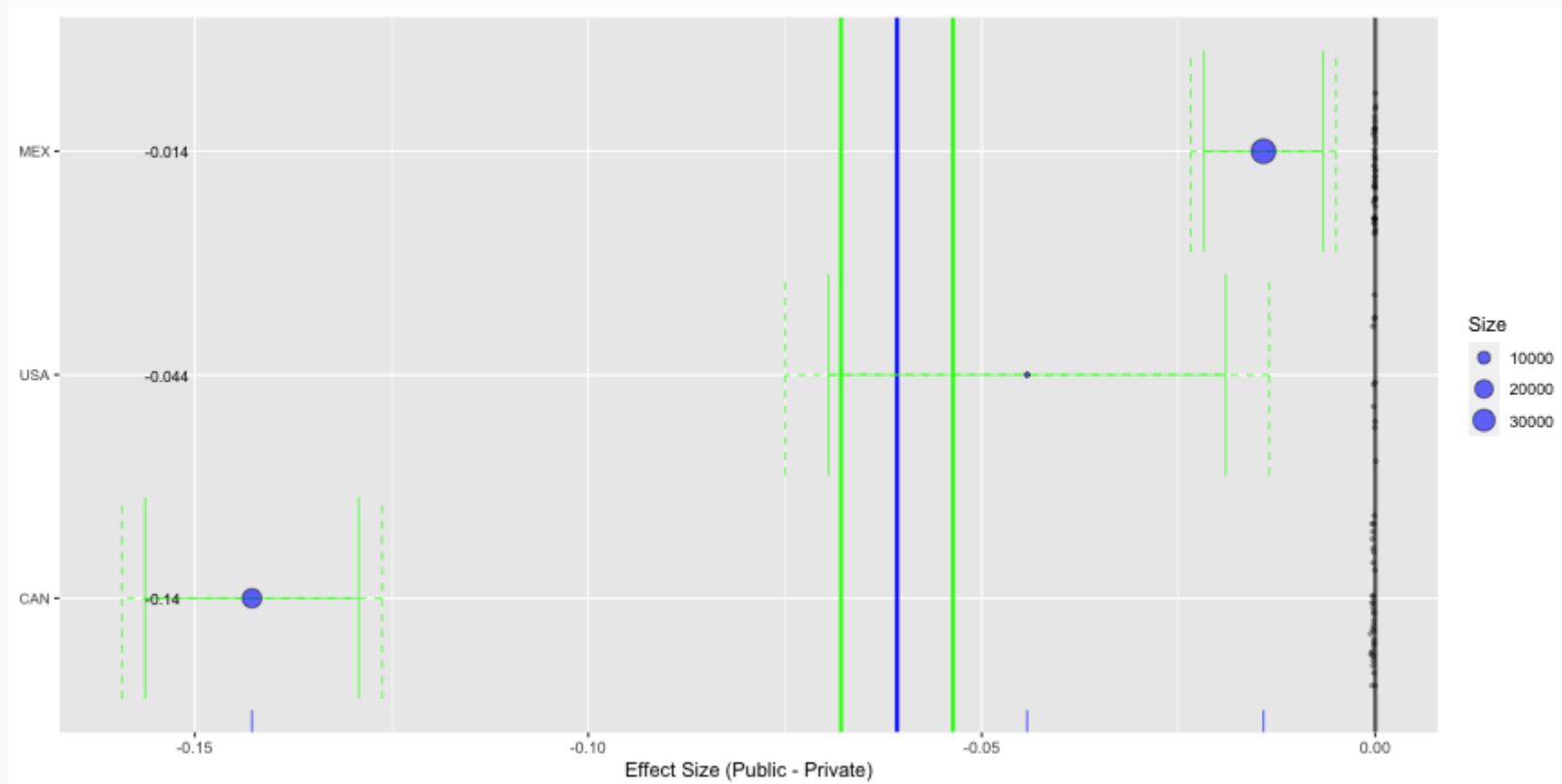


Multilevel PSA Assessment Plot

The multilevel PSA assessment plot is an extension of the `circ.psa` plot in `PSAgraphics` introduced by Helmreich and Pruzek (2009).

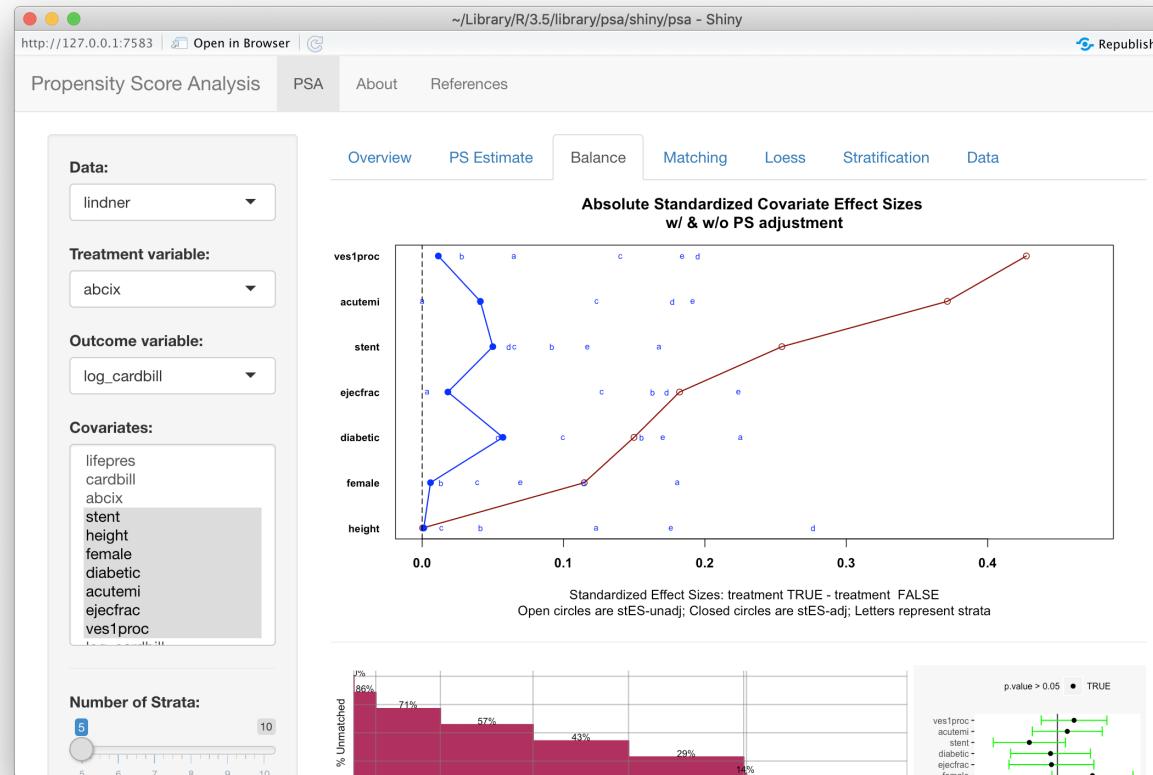
```
plot(results.psa.math)
```

Multilevel PSA Difference Plot



Shiny Application

psa::psa_shiny()



Questions

One Minute Paper

Complete the one minute paper:

<https://forms.gle/yB3ds6MYE89Z1pURA>

1. What was the most important thing you learned during this class?
2. What important question remains unanswered for you?