Causal Inference

Seminar #1

Juraj Medzihorsky

SGIA & RMC
Durham University

16 January 2024

Matching

- Suppose

$$D_i \longrightarrow Y_i$$
$$Y_i(0), Y_i(1) \not\perp D_i$$

that is, there is selection into treatment, and also

$$\boldsymbol{X}_i \longrightarrow D_i$$
$$\boldsymbol{X}_i \longrightarrow Y_i$$

- Intuitively: If we are able to account for all of the $\boldsymbol{X}$ variables that affect both selection into treatment, D, **and** variation in the outcome, Y, then we can claim independence between D and potential outcomes.

- Suppose

$$D_i \longrightarrow Y_i$$
$$Y_i(0), Y_i(1) \not\perp D_i$$

that is, there is selection into treatment, and also

$$\boldsymbol{X}_i \longrightarrow D_i$$
$$\boldsymbol{X}_i \longrightarrow Y_i$$

- Intuitively: If we are able to account for all of the $\boldsymbol{X}$ variables that affect both selection into treatment, D, **and** variation in the outcome, Y, then we can claim independence between D and potential outcomes.
- The big problem: We have to be able to observe and measure all of the X variables that affect selection into treatment and variation in outcomes.

- Suppose

$$\begin{aligned} D_i &\longrightarrow Y_i \\ Y_i(0), Y_i(1) &\not\perp D_i \end{aligned}$$

that is, there is selection into treatment, and also

$$\begin{aligned} \boldsymbol{X}_i &\longrightarrow D_i \\ \boldsymbol{X}_i &\longrightarrow Y_i \end{aligned}$$

- Intuitively: If we are able to account for all of the $\boldsymbol{X}$ variables that affect both selection into treatment, D, **and** variation in the outcome, Y, then we can claim independence between D and potential outcomes.
- The big problem: We have to be able to observe and measure all of the X variables that affect selection into treatment and variation in outcomes.

Matching

- More precisely *matching on observables*
- Suppose
  - Binary treatment, i.e., only two options: treatment and control
  - We know and observe all potential confounders
- Then we can use the untreated observations to construct a control group that will be as similar as possible to the treated observations
- This helps in estimating the average treatment effect *on the treated* (ATT) which may not be the same as the average treatment effect (ATE)
- Many options: different matching algorithms

Identification Assumptions

1. *Selection on Observables*

Identification Assumptions

1. *Selection on Observables*
    - *ATE Version:* $(Y(1), Y(0)) \perp\!\!\!\perp D | X$

Identification Assumptions

1. *Selection on Observables*
   - *ATE Version:* $(Y(1), Y(0)) \perp\!\!\!\perp D|X$
     - *There exists a set of observable covariates, X, such that after controlling for X, potential outcomes are independent of treatment status*

Identification Assumptions

1. *Selection on Observables*
    - *ATE Version:* $(Y(1), Y(0)) \perp\!\!\!\perp D | X$
        - *There exists a set of observable covariates, X, such that after controlling for X, potential outcomes are independent of treatment status*
    - *ATT Version:* $Y(0) \perp\!\!\!\perp D | X$

Identification Assumptions

1. *Selection on Observables*
   - *ATE Version:* $(Y(1), Y(0)) \perp\!\!\!\perp D | X$
     - *There exists a set of observable covariates, X, such that after controlling for X, potential outcomes are independent of treatment status*
   - *ATT Version:* $Y(0) \perp\!\!\!\perp D | X$
     - *There exists a set of observable covariates, X, such that after controlling for X, counterfactual outcomes for treated units and observed outcomes for untreated units are independent of treatment status*

Identification Assumptions

1. *Selection on Observables*
    - *ATE Version:* $(Y(1), Y(0)) \perp\!\!\!\perp D | X$
        - *There exists a set of observable covariates, X, such that after controlling for X, potential outcomes are independent of treatment status*
    - *ATT Version:* $Y(0) \perp\!\!\!\perp D | X$
        - *There exists a set of observable covariates, X, such that after controlling for X, counterfactual outcomes for treated units and observed outcomes for untreated units are independent of treatment status*
2. *Common Support*

Identification Assumptions

1. *Selection on Observables*
   - *ATE Version:* $(Y(1), Y(0)) \perp\!\!\!\perp D | X$
     - *There exists a set of observable covariates, X, such that after controlling for X, potential outcomes are independent of treatment status*
   - *ATT Version:* $Y(0) \perp\!\!\!\perp D | X$
     - *There exists a set of observable covariates, X, such that after controlling for X, counterfactual outcomes for treated units and observed outcomes for untreated units are independent of treatment status*
2. *Common Support*
   - *ATE Version:* $0 < \Pr(D = 1 | X = x) < 1$

Identification Assumptions

1. *Selection on Observables*
   - *ATE Version:* $(Y(1), Y(0)) \perp\!\!\!\perp D | X$
     - *There exists a set of observable covariates, X, such that after controlling for X, potential outcomes are independent of treatment status*
   - *ATT Version:* $Y(0) \perp\!\!\!\perp D | X$
     - *There exists a set of observable covariates, X, such that after controlling for X, counterfactual outcomes for treated units and observed outcomes for untreated units are independent of treatment status*
2. *Common Support*
   - *ATE Version:* $0 < \Pr(D = 1 | X = x) < 1$
     - *For each value of X, there is a positive probability of observing both treated and untreated units*

Identification Assumptions

1. *Selection on Observables*
   - *ATE Version:* $(Y(1), Y(0)) \perp\!\!\!\perp D|X$
     - *There exists a set of observable covariates, X, such that after controlling for X, potential outcomes are independent of treatment status*
   - *ATT Version:* $Y(0) \perp\!\!\!\perp D|X$
     - *There exists a set of observable covariates, X, such that after controlling for X, counterfactual outcomes for treated units and observed outcomes for untreated units are independent of treatment status*
2. *Common Support*
   - *ATE Version:* $0 < \Pr(D = 1|X = x) < 1$
     - *For each value of X, there is a positive probability of observing both treated and untreated units*
   - *ATT Version:* $\Pr(D = 1|X = x) < 1$, *with* $\Pr(D = 1) > 0$

Identification Assumptions

1. *Selection on Observables*
   - *ATE Version:* $(Y(1), Y(0)) \perp\!\!\!\perp D|X$
     - *There exists a set of observable covariates, X, such that after controlling for X, potential outcomes are independent of treatment status*
   - *ATT Version:* $Y(0) \perp\!\!\!\perp D|X$
     - *There exists a set of observable covariates, X, such that after controlling for X, counterfactual outcomes for treated units and observed outcomes for untreated units are independent of treatment status*
2. *Common Support*
   - *ATE Version:* $0 < \Pr(D = 1|X = x) < 1$
     - *For each value of X, there is a positive probability of observing both treated and untreated units*
   - *ATT Version:* $\Pr(D = 1|X = x) < 1$, *with* $\Pr(D = 1) > 0$
     - *For each value of X observed for a treated unit, there should exist an untreated unit with the same value of X*

When X is continuous we can estimate α_{ATT} by "imputing" the missing potential outcome of each treated unit using the observed outcome from the "closest" control unit:

$$\widehat{\alpha}_{ATT} = \frac{1}{N_1} \sum_{D_i=1} \left(Y_i - Y_{j(i)} \right)$$

where $Y_{j(i)}$ is the outcome of an untreated observation such that $X_{j(i)}$ is the closest value to X_i among the untreated observations.

When X is continuous we can estimate α_{ATT} by "imputing" the missing potential outcome of each treated unit using the observed outcome from the "closest" control unit:

$$\widehat{\alpha}_{ATT} = \frac{1}{N_1} \sum_{D_i=1} \left(Y_i - Y_{j(i)}\right)$$

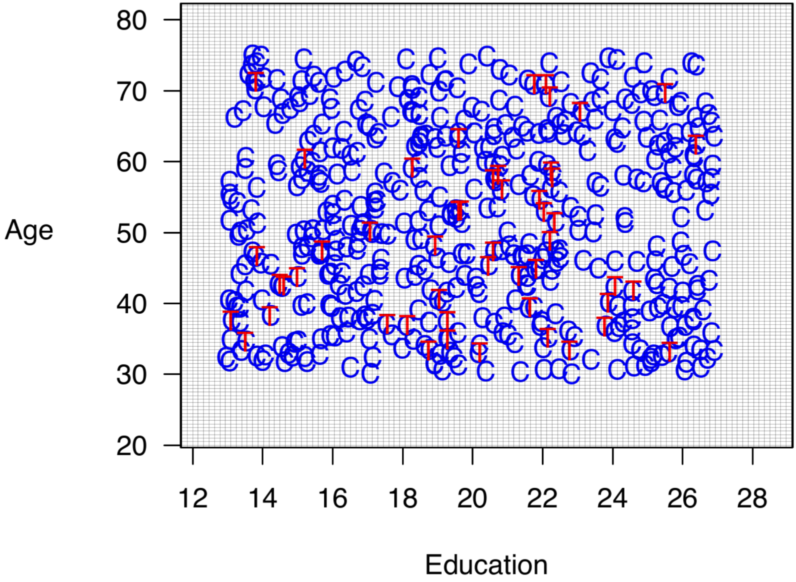
where $Y_{j(i)}$ is the outcome of an untreated observation such that $X_{j(i)}$ is the closest value to X_i among the untreated observations.

We can also use the average for M closest matches:

$$\widehat{\alpha}_{ATT} = \frac{1}{N_1} \sum_{D_i=1} \left\{ Y_i - \left(\frac{1}{M} \sum_{m=1}^{M} Y_{j_{m(i)}} \right) \right\}$$

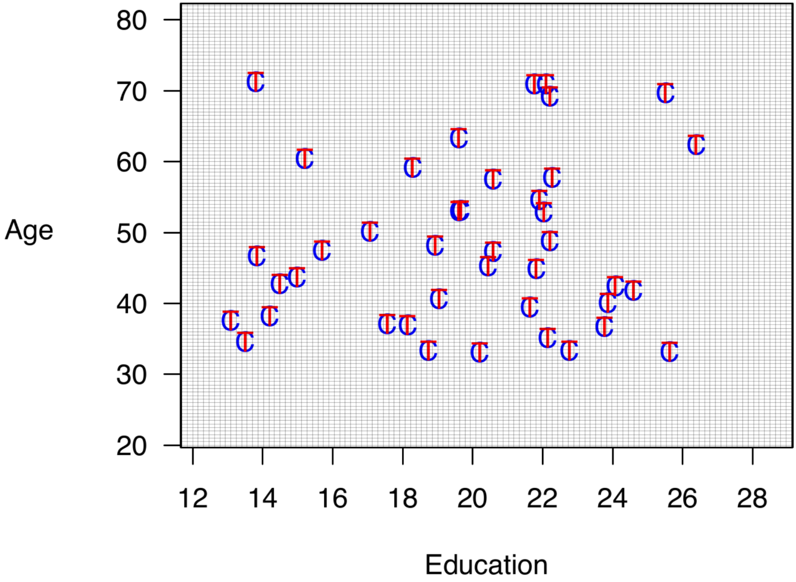
Works well when we can find good matches for each treated unit.

Exact Matching: 2D Example



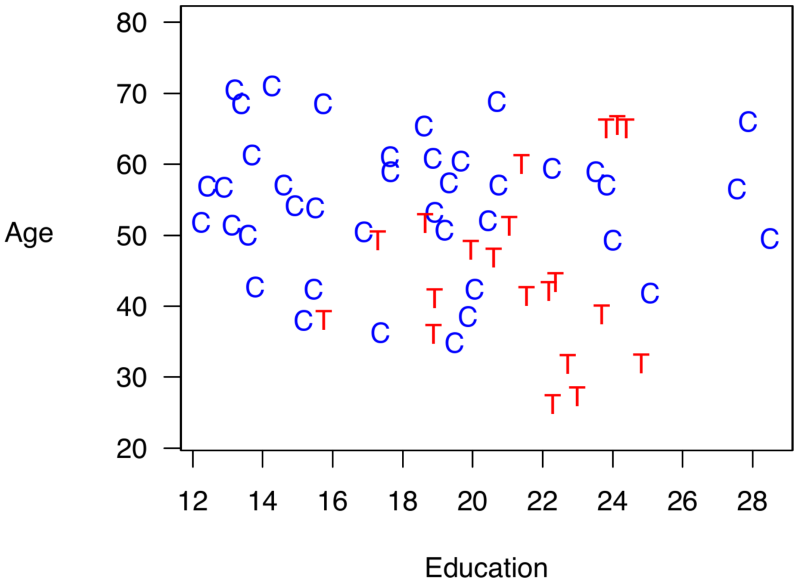
Source: Gary King

Exact Matching: 2D Example



Source: Gary King

Coarsened Exact Matching: Example



Source: Gary King

Matching: Distance Metric

"Closeness" is often defined by a distance metric that projects the distance between the multivariate covariate vectors $X_i = (X_{i1}, X_{i2}, ..., X_{ik})'$ and $X_j = (X_{j1}, X_{j2}, ..., X_{jk})'$ onto a univariate scale

Matching: Distance Metric

"Closeness" is often defined by a distance metric that projects the distance between the multivariate covariate vectors $X_i = (X_{i1}, X_{i2}, ..., X_{ik})'$ and $X_j = (X_{j1}, X_{j2}, ..., X_{jk})'$ onto a univariate scale

A commonly used distance is the Mahalanobis distance:

$$D_M(X_i, X_j) = \sqrt{(X_i - X_j)' S^{-1} (X_i - X_j)} \ ,$$

where S is the sample variance-covariance-matrix.

Matching: Distance Metric

"Closeness" is often defined by a distance metric that projects the distance between the multivariate covariate vectors $X_i = (X_{i1}, X_{i2}, ..., X_{ik})'$ and $X_j = (X_{j1}, X_{j2}, ..., X_{jk})'$ onto a univariate scale

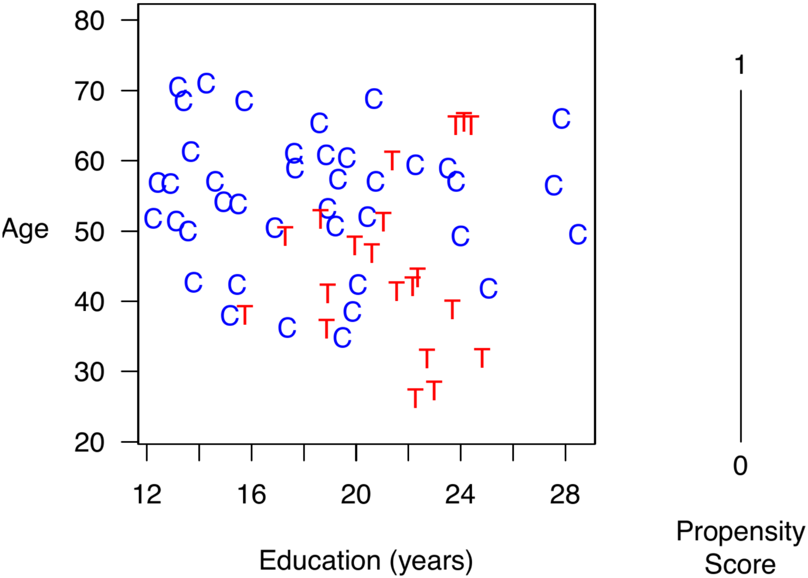
A commonly used distance is the Mahalanobis distance:

$$D_M(X_i, X_j) = \sqrt{(X_i - X_j)' S^{-1} (X_i - X_j)} \ ,$$

where S is the sample variance-covariance-matrix.

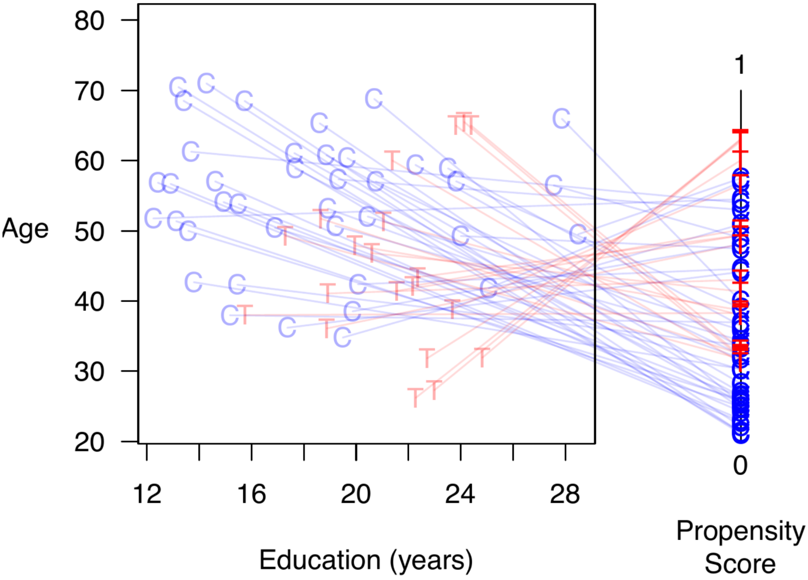
Propensity score is defined as the selection probability conditional on the confounding variables: $\pi(X) = P(D = 1|X)$.

Propensity Score Matching: Example



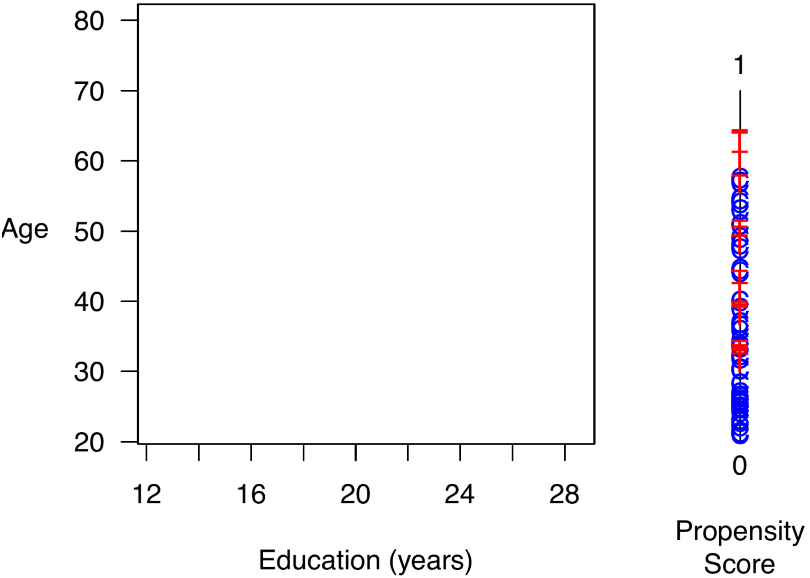
Source: Gary King

Propensity Score Matching: Example



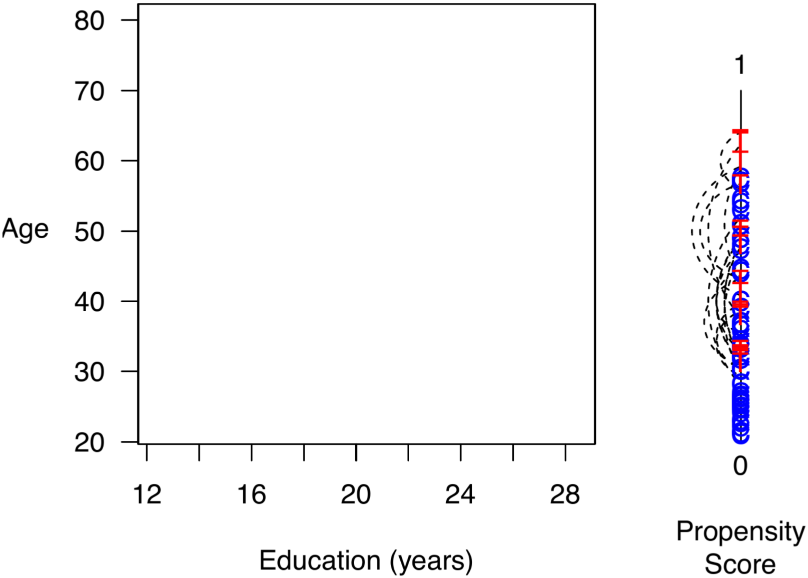
Source: Gary King

Propensity Score Matching: Example



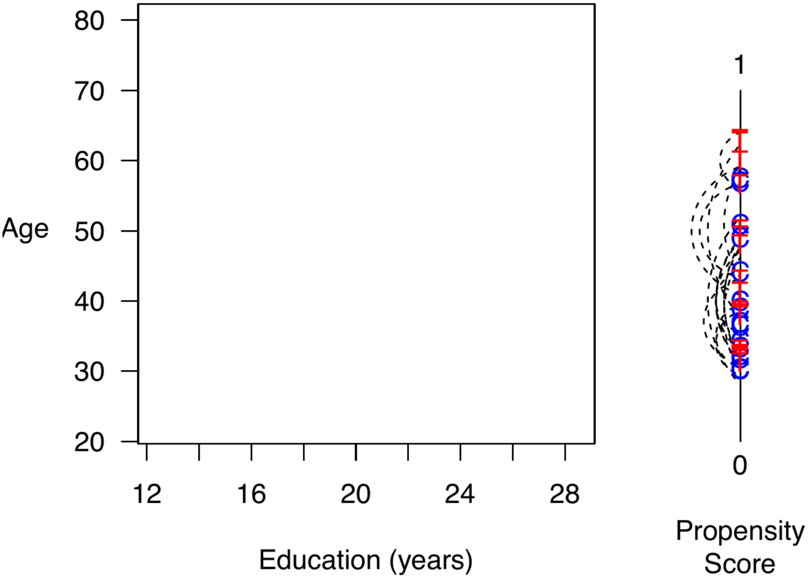
Source: Gary King

Propensity Score Matching: Example



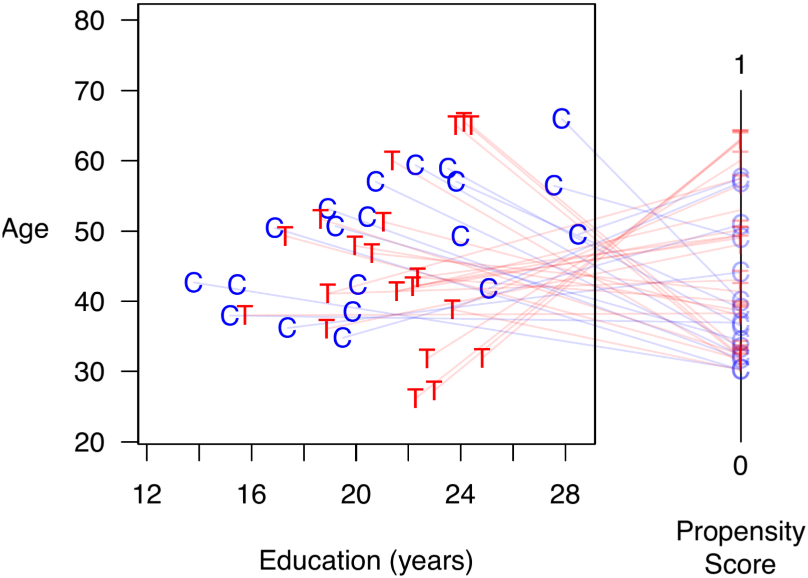
Source: Gary King

Propensity Score Matching: Example



Source: Gary King

Propensity Score Matching: Example



Source: Gary King

Practical Steps in Matching

1. Like all data analysis exercises, explore your raw data before making a lot of assumptions, transformations, etc.

1. Like all data analysis exercises, explore your raw data before making a lot of assumptions, transformations, etc.

   In matching exercises, prior to constructing a matched sample, this can involve...

   - Simple difference-of-means tests on the outcome variable
   - Regression-adjusted difference-of-means tests on the outcome variable using pre-treatment covariates

Practical Steps in Matching

2. Construct the matched sample
    - Whether using distance-matching or propensity scores, try multiple specifications, using different choices for covariates, numbers of control units for each treatment unit, matching with and without replacement, etc.
    - Check for balance on pre-treatment covariates (e.g., t-tests for balance on pre-treatment covariates, graphical overlays of covariate densities between treatment groups)
    - For propensity-score matching, explore the distribution of the propensity score across treatment groups

2. Construct the matched sample

    - Whether using distance-matching or propensity scores, try multiple specifications, using different choices for covariates, numbers of control units for each treatment unit, matching with and without replacement, etc.
    - Check for balance on pre-treatment covariates (e.g., t-tests for balance on pre-treatment covariates, graphical overlays of covariate densities between treatment groups)
    - For propensity-score matching, explore the distribution of the propensity score across treatment groups
    - Choose your method based on balance, not treatment effects!

      For best practices, do not even look at treatment effects before the matching method is chosen

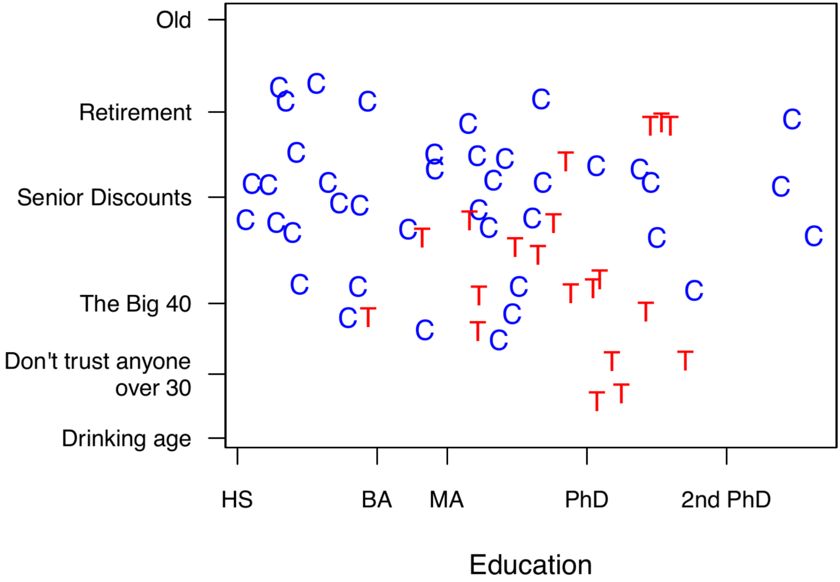
3. Estimate treatment effects
    - Simple difference-of-means tests between experimental groups
    - Regression-adjusted difference of means tests using pre-treatment variables as controls
    - Same types of graphical techniques used for difference-of-means and regressions
    - Comparisons to unmatched analyses
    - If possible, comparisons to an experimental benchmark

4. Evaluate the selection-on-observables assumption with sensitivity analysis
    - Rosenbaum bounds are probably the most widely used test
    - Usually will take the form of asking: How bad will our inferences be if we left out an important covariate?

?

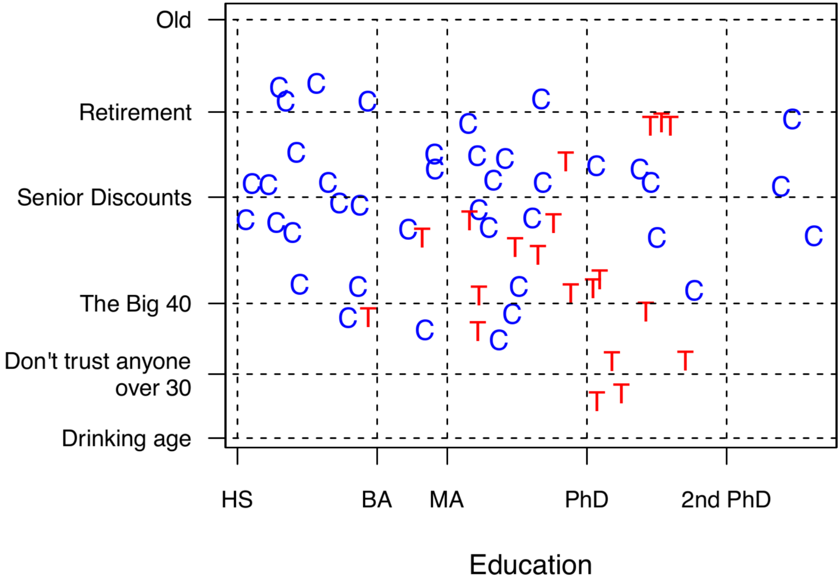
Thank you!

Coarsened Exact Matching: Example



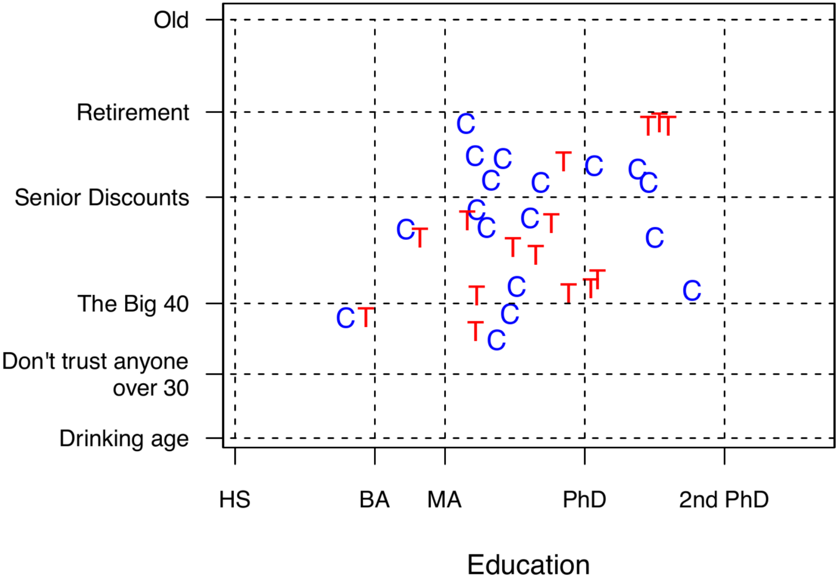
Source: Gary King

Coarsened Exact Matching: Example



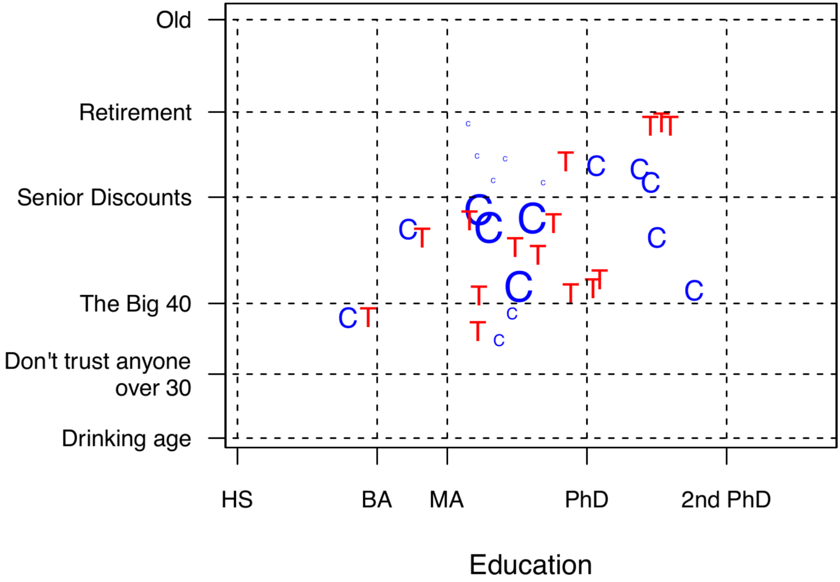
Source: Gary King

Coarsened Exact Matching: Example



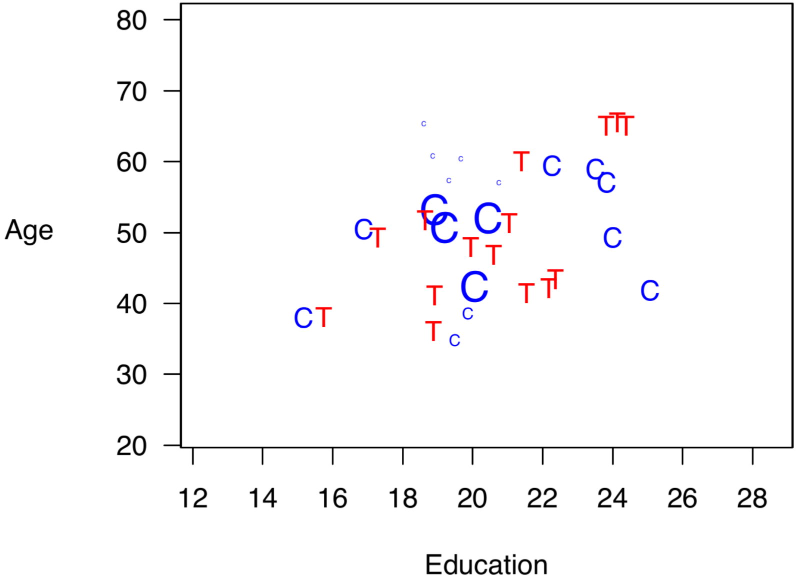
Source: Gary King

Coarsened Exact Matching: Example



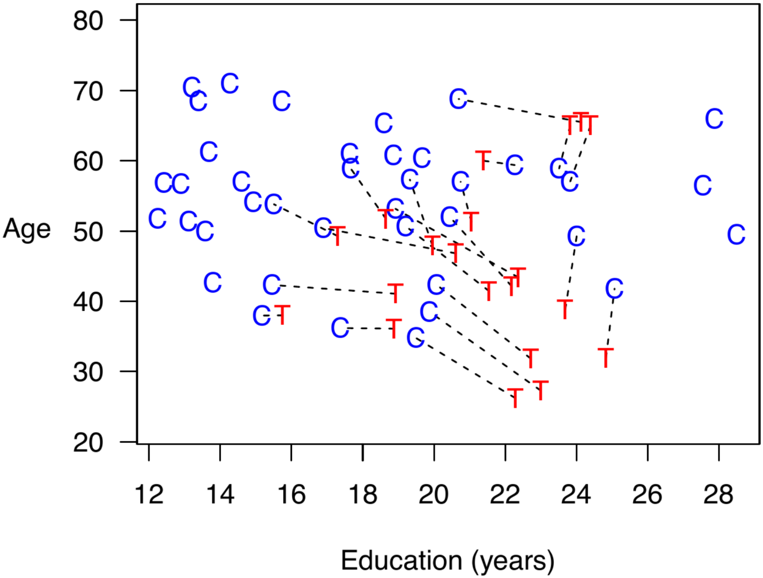
Source: Gary King

Coarsened Exact Matching: Example



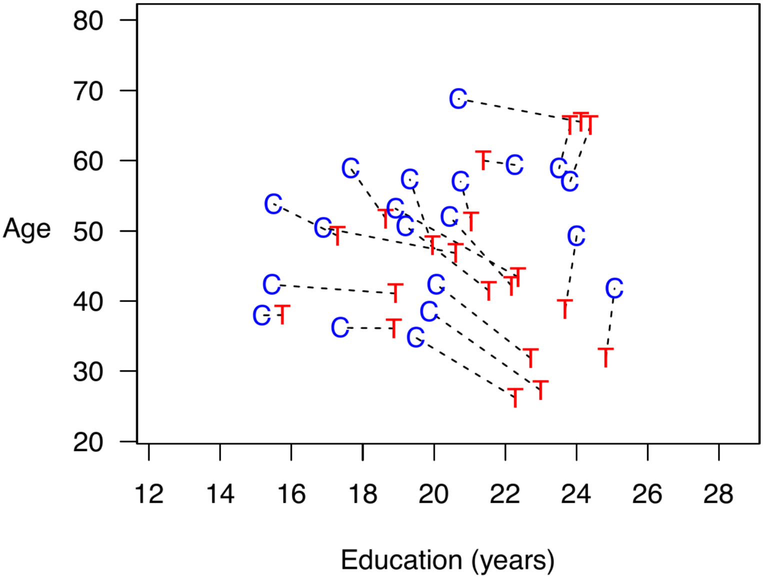
Source: Gary King

Mahalanobis Distance Matching: Example



Source: Gary King

Mahalanobis Distance Matching: Example



Source: Gary King