Advanced Quantitative Research Methodology, Lecture Notes: Matching Methods for Causal Inference[1]

Gary King[2]

Institute for Quantitative Social Science
Harvard University

[2] GaryKing.org.

Overview

Overview

- Problem: Model dependence (which we've seen)

Overview

- Problem: Model dependence (which we've seen)
- Solution: Matching to reduce model dependence

Overview

- Problem: Model dependence (which we've seen)
- Solution: Matching to reduce model dependence
- Problem: Matching prunes n to improve imbalance, but

Overview

- Problem: Model dependence (which we've seen)
- Solution: Matching to reduce model dependence
- Problem: Matching prunes n to improve imbalance, but
    - Some: set n and don't guarantee imbalance

Overview

- Problem: Model dependence (which we've seen)
- Solution: Matching to reduce model dependence
- Problem: Matching prunes n to improve imbalance, but
  - Some: set n and don't guarantee imbalance
  - Others: set imbalance and don't guarantee n

Overview

- Problem: Model dependence (which we've seen)
- Solution: Matching to reduce model dependence
- Problem: Matching prunes n to improve imbalance, but
  - Some: set n and don't guarantee imbalance
  - Others: set imbalance and don't guarantee n
  - Plus: Matching methods optimize a different "imbalance" than recommended post-hoc checks

Overview

- Problem: Model dependence (which we've seen)
- Solution: Matching to reduce model dependence
- Problem: Matching prunes n to improve imbalance, but
  - Some: set n and don't guarantee imbalance
  - Others: set imbalance and don't guarantee n
  - Plus: Matching methods optimize a different "imbalance" than recommended post-hoc checks
- Solution: easier & more powerful

Overview

- Problem: Model dependence (which we've seen)
- Solution: Matching to reduce model dependence
- Problem: Matching prunes n to improve imbalance, but
  - Some: set n and don't guarantee imbalance
  - Others: set imbalance and don't guarantee n
  - Plus: Matching methods optimize a different "imbalance" than recommended post-hoc checks
- Solution: easier & more powerful
  - Estimate the (n-imbalance) "matching frontier"

Overview

- Problem: Model dependence (which we've seen)
- Solution: Matching to reduce model dependence
- Problem: Matching prunes n to improve imbalance, but
  - Some: set n and don't guarantee imbalance
  - Others: set imbalance and don't guarantee n
  - Plus: Matching methods optimize a different "imbalance" than recommended post-hoc checks
- Solution: easier & more powerful
  - Estimate the (n-imbalance) "matching frontier"
  - Imbalance metric choice defines the frontier

Overview

- Problem: Model dependence (which we've seen)
- Solution: Matching to reduce model dependence
- Problem: Matching prunes n to improve imbalance, but
  - Some: set n and don't guarantee imbalance
  - Others: set imbalance and don't guarantee n
  - Plus: Matching methods optimize a different "imbalance" than recommended post-hoc checks
- Solution: easier & more powerful
  - Estimate the (n-imbalance) "matching frontier"
  - Imbalance metric choice defines the frontier
- Side point:

Overview

- Problem: Model dependence (which we've seen)
- Solution: Matching to reduce model dependence
- Problem: Matching prunes n to improve imbalance, but
  - Some: set n and don't guarantee imbalance
  - Others: set imbalance and don't guarantee n
  - Plus: Matching methods optimize a different "imbalance" than recommended post-hoc checks
- Solution: easier & more powerful
  - Estimate the (n-imbalance) "matching frontier"
  - Imbalance metric choice defines the frontier
- Side point:
  - Problem: Propensity score matching increases imbalance!

Overview

- Problem: Model dependence (which we've seen)
- Solution: Matching to reduce model dependence
- Problem: Matching prunes n to improve imbalance, but
  - Some: set n and don't guarantee imbalance
  - Others: set imbalance and don't guarantee n
  - Plus: Matching methods optimize a different "imbalance" than recommended post-hoc checks
- Solution: easier & more powerful
  - Estimate the (n-imbalance) "matching frontier"
  - Imbalance metric choice defines the frontier
- Side point:
  - Problem: Propensity score matching increases imbalance!
  - Solution: Not an issue with other methods

Overview

- Problem: Model dependence (which we've seen)
- Solution: Matching to reduce model dependence
- Problem: Matching prunes n to improve imbalance, but
  - Some: set n and don't guarantee imbalance
  - Others: set imbalance and don't guarantee n
  - Plus: Matching methods optimize a different "imbalance" than recommended post-hoc checks
- Solution: easier & more powerful
  - Estimate the (n-imbalance) "matching frontier"
  - Imbalance metric choice defines the frontier
- Side point:
  - Problem: Propensity score matching increases imbalance!
  - Solution: Not an issue with other methods
- Formal Assumptions: We save for last

Overview of Matching for Causal Inference

Overview of Matching for Causal Inference

- Goal: reduce model dependence

Overview of Matching for Causal Inference

- Goal: reduce model dependence
- A nonparametric, non-model-based approach

Overview of Matching for Causal Inference

- Goal: reduce model dependence
- A nonparametric, non-model-based approach
- Makes parametric models work better rather than substitute for them (i.e,. matching is not an estimator; its a preprocessing method)

Overview of Matching for Causal Inference

- Goal: reduce model dependence
- A nonparametric, non-model-based approach
- Makes parametric models work better rather than substitute for them (i.e,. matching is not an estimator; its a preprocessing method)
- Should have been called pruning (no bias is introduced if pruning is a function of T and X, but not Y)

Overview of Matching for Causal Inference

- Goal: reduce model dependence
- A nonparametric, non-model-based approach
- Makes parametric models work better rather than substitute for them (i.e,. matching is not an estimator; its a preprocessing method)
- Should have been called pruning (no bias is introduced if pruning is a function of T and X, but not Y)
- Apply model to preprocessed (pruned) rather than raw data

Overview of Matching for Causal Inference

- Goal: reduce model dependence
- A nonparametric, non-model-based approach
- Makes parametric models work better rather than substitute for them (i.e,. matching is not an estimator; its a preprocessing method)
- Should have been called pruning (no bias is introduced if pruning is a function of T and X, but not Y)
- Apply model to preprocessed (pruned) rather than raw data
- Violates the "more data is better" principle, but that only applies when you know the DGP

Overview of Matching for Causal Inference

- Goal: reduce model dependence
- A nonparametric, non-model-based approach
- Makes parametric models work better rather than substitute for them (i.e,. matching is not an estimator; its a preprocessing method)
- Should have been called pruning (no bias is introduced if pruning is a function of T and X, but not Y)
- Apply model to preprocessed (pruned) rather than raw data
- Violates the "more data is better" principle, but that only applies when you know the DGP
- Overall idea:

Overview of Matching for Causal Inference

- Goal: reduce model dependence
- A nonparametric, non-model-based approach
- Makes parametric models work better rather than substitute for them (i.e,. matching is not an estimator; its a preprocessing method)
- Should have been called pruning (no bias is introduced if pruning is a function of T and X, but not Y)
- Apply model to preprocessed (pruned) rather than raw data
- Violates the "more data is better" principle, but that only applies when you know the DGP
- Overall idea:
  - If each treated unit exactly matches a control unit w.r.t. X, then: (1) treated and control groups are identical, (2) X is no longer a confounder, (3) no need to worry about the functional form ($\bar{Y}_T - \bar{Y}_C$ is good enough).

Overview of Matching for Causal Inference

- Goal: reduce model dependence
- A nonparametric, non-model-based approach
- Makes parametric models work better rather than substitute for them (i.e,. matching is not an estimator; its a preprocessing method)
- Should have been called pruning (no bias is introduced if pruning is a function of T and X, but not Y)
- Apply model to preprocessed (pruned) rather than raw data
- Violates the "more data is better" principle, but that only applies when you know the DGP
- Overall idea:
  - If each treated unit exactly matches a control unit w.r.t. X, then: (1) treated and control groups are identical, (2) X is no longer a confounder, (3) no need to worry about the functional form ($\bar{Y}_T - \bar{Y}_C$ is good enough).
  - If treated and control groups are better balanced than when you started, due to pruning, model dependence is reduced

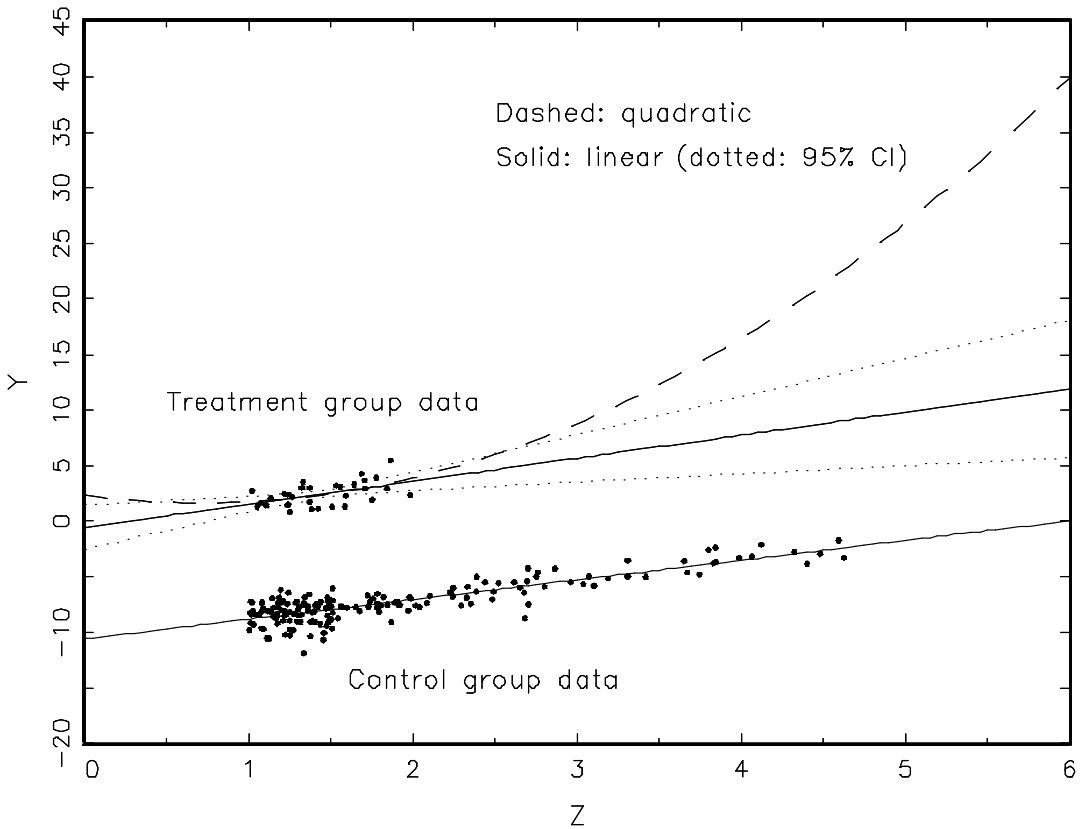
Model Dependence: A Simpler Example

Model Dependence: A Simpler Example

(King and Zeng, 2006: fig.4 *Political Analysis*)

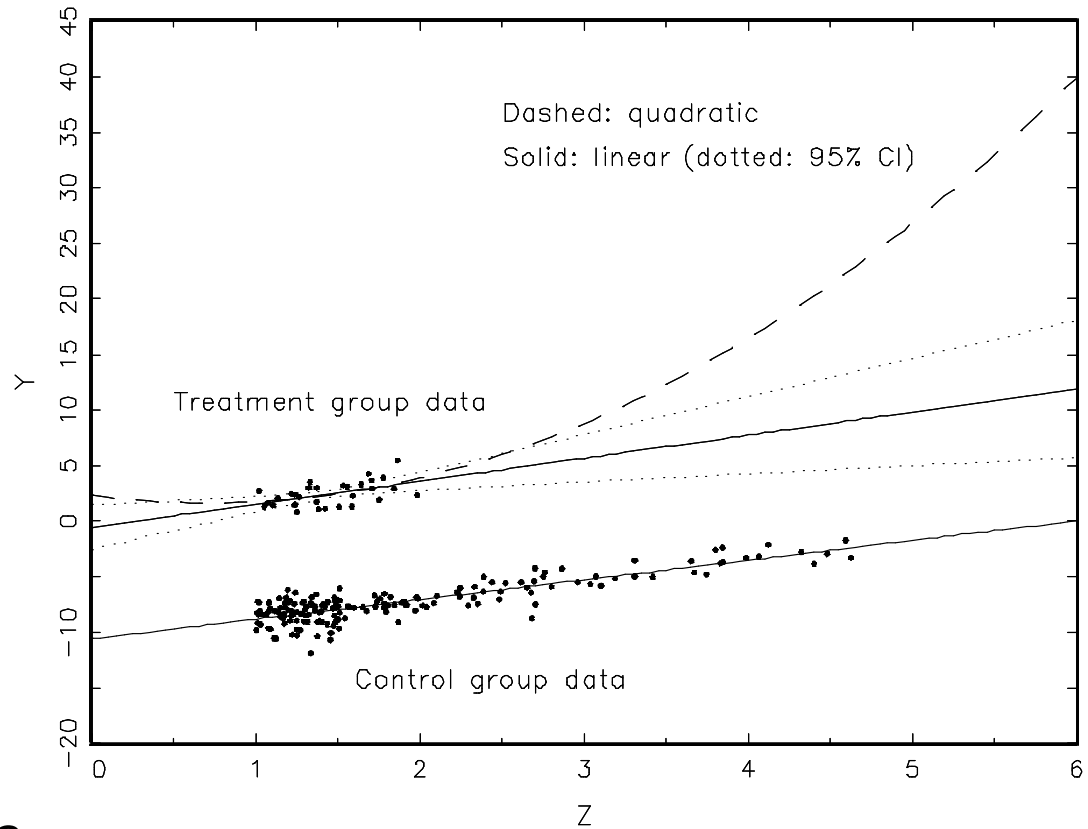
Model Dependence: A Simpler Example

(King and Zeng, 2006: fig.4 *Political Analysis*)



Model Dependence: A Simpler Example

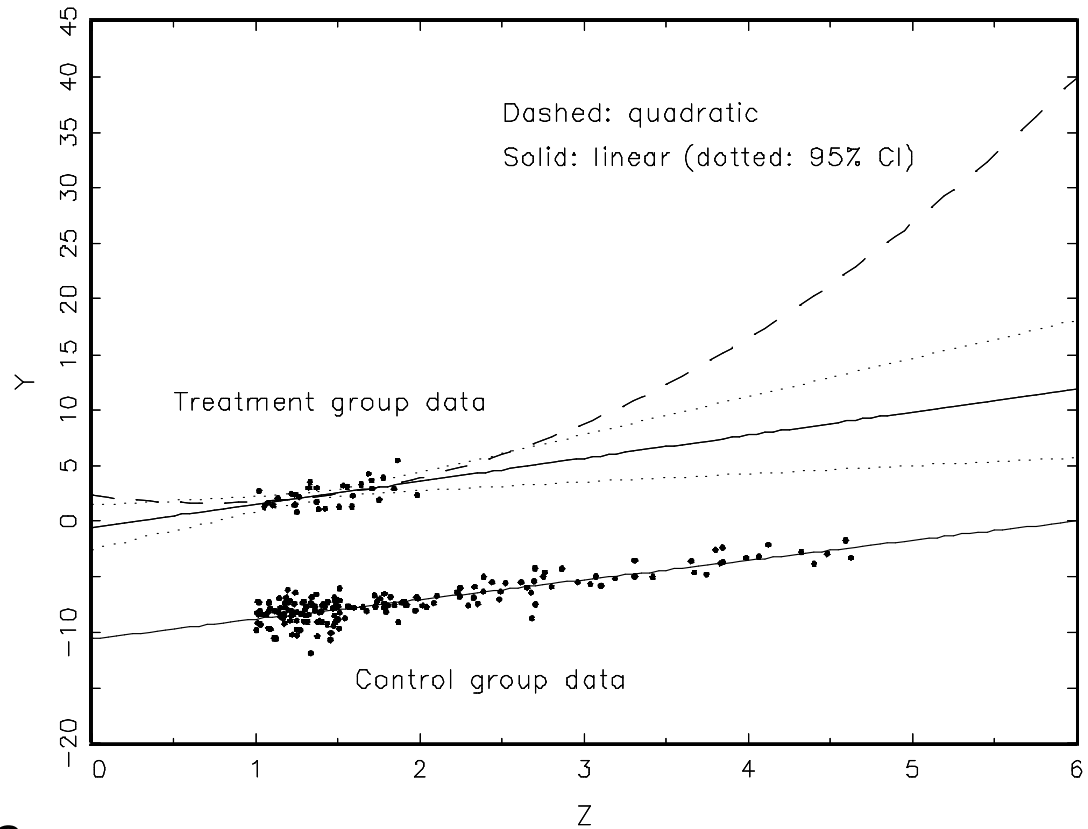
(King and Zeng, 2006: fig.4 *Political Analysis*)



What to do?

Model Dependence: A Simpler Example

(King and Zeng, 2006: fig.4 *Political Analysis*)

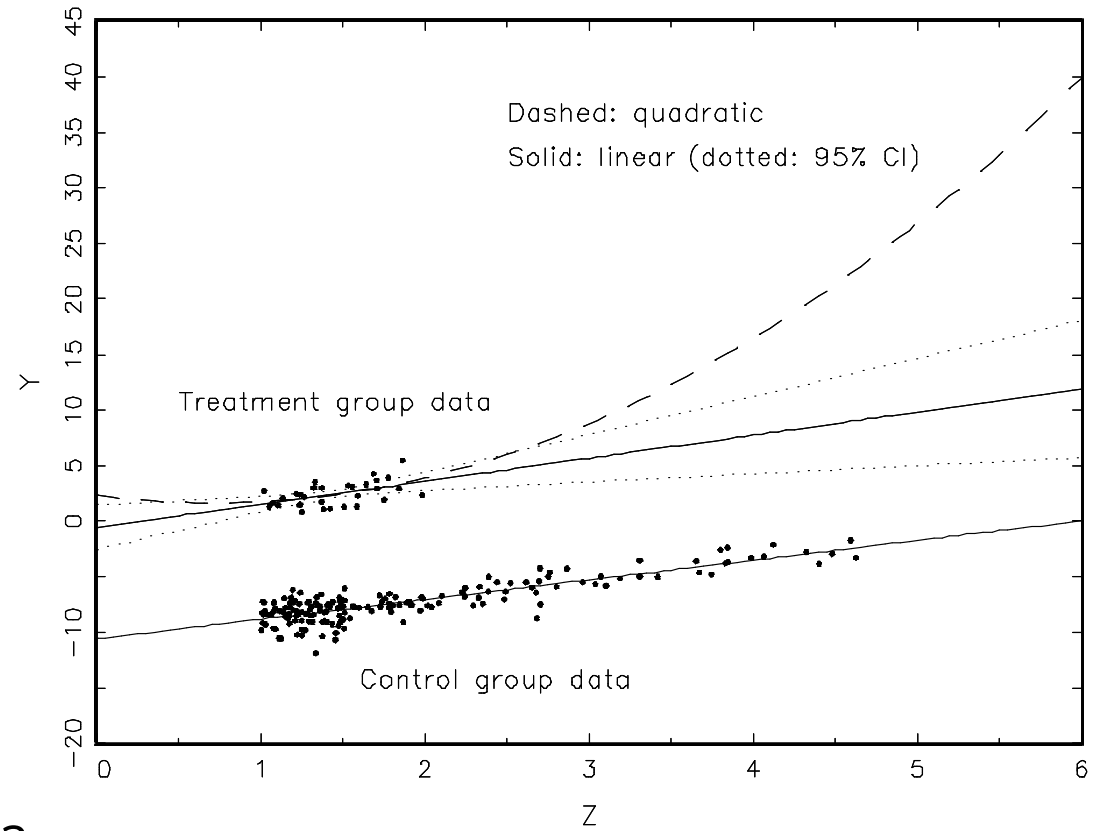


What to do?

- Preprocess I: Eliminate extrapolation region

Model Dependence: A Simpler Example

(King and Zeng, 2006: fig.4 *Political Analysis*)

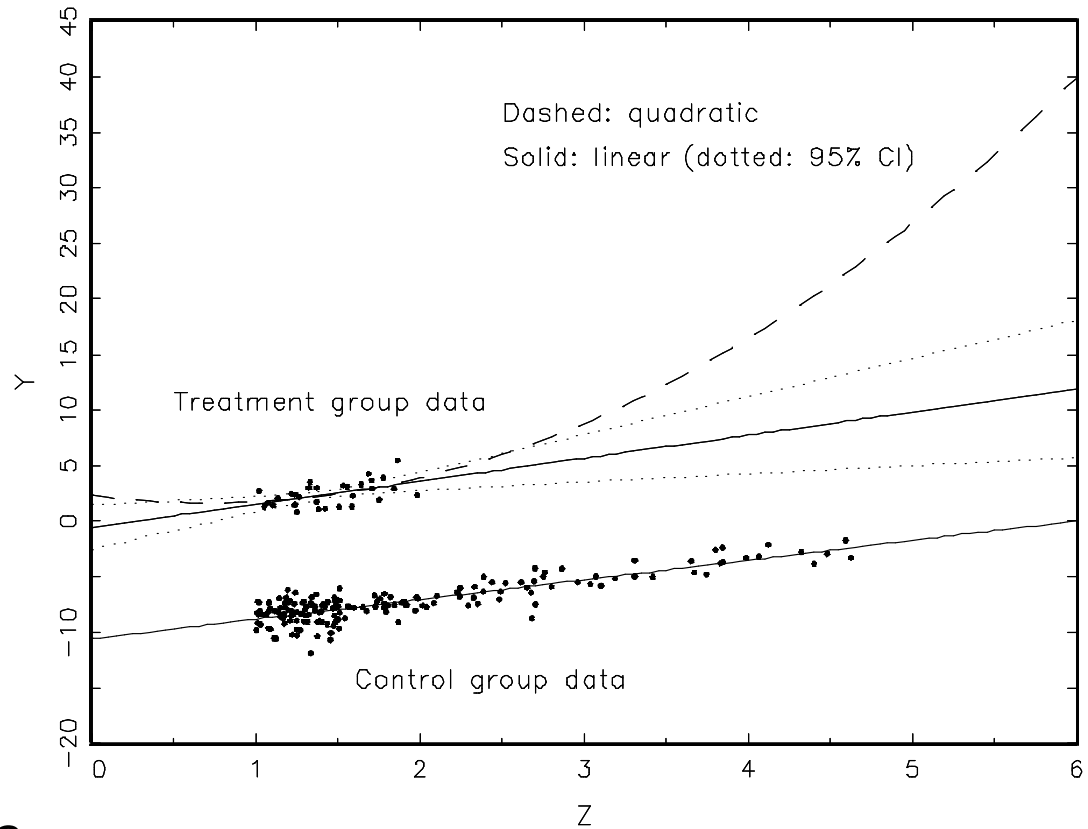


What to do?

- Preprocess I: Eliminate extrapolation region
- Preprocess II: Match (prune) within interpolation region

Model Dependence: A Simpler Example

(King and Zeng, 2006: fig.4 *Political Analysis*)

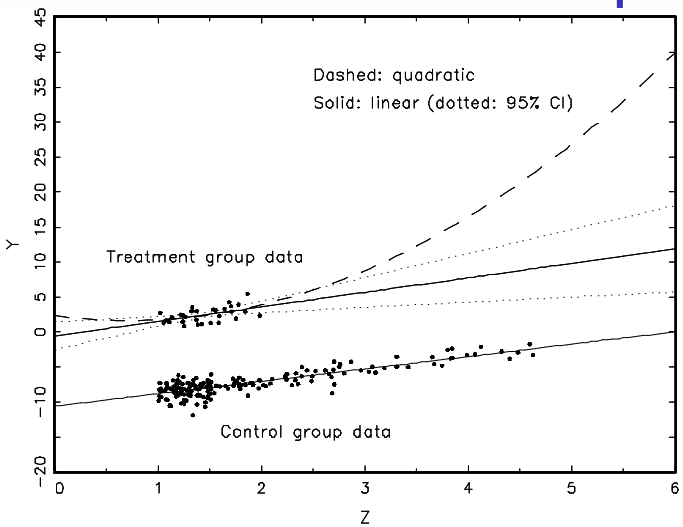


What to do?

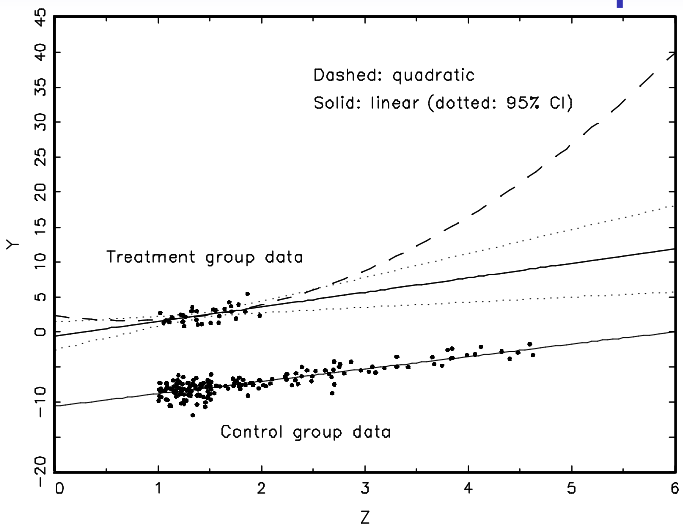
- Preprocess I: Eliminate extrapolation region
- Preprocess II: Match (prune) within interpolation region
- Model remaining imbalance (as you would w/o matching)

Remove Extrapolation Region, then Match

Remove Extrapolation Region, then Match

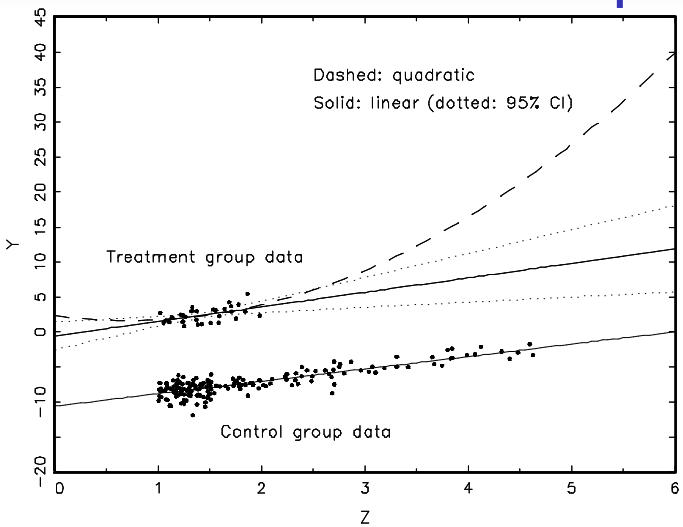


Remove Extrapolation Region, then Match



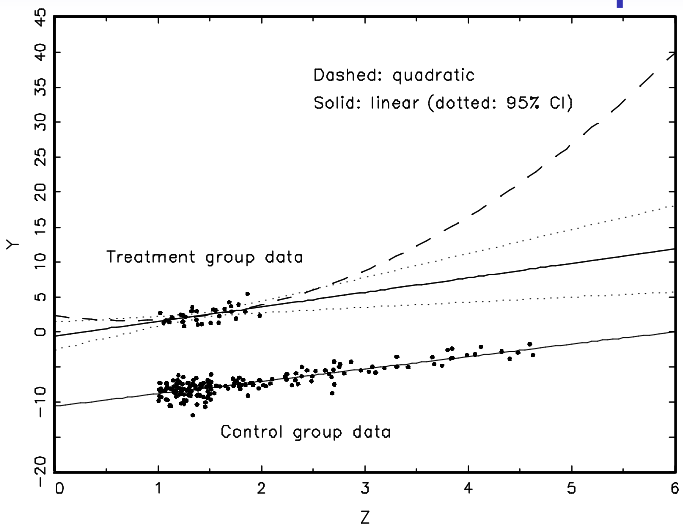
- Must remove data (selecting on X) to avoid extrapolation.

Remove Extrapolation Region, then Match



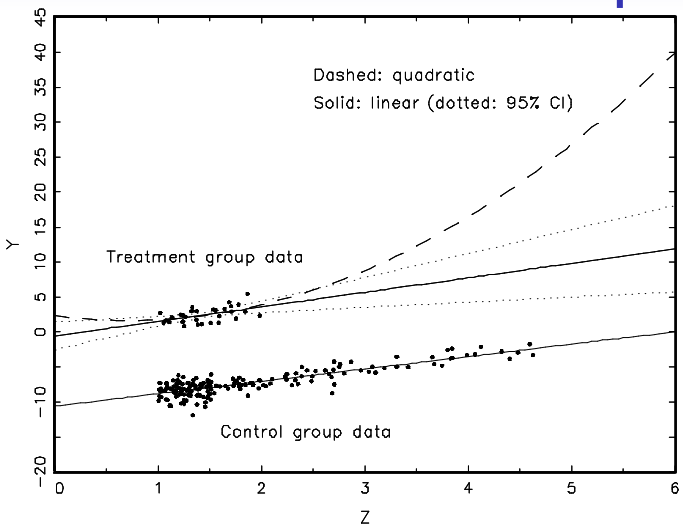
- Must remove data (selecting on X) to avoid extrapolation.
- Options to find "common support" of $p(X|T=1)$ and $P(X|T=0)$

Remove Extrapolation Region, then Match



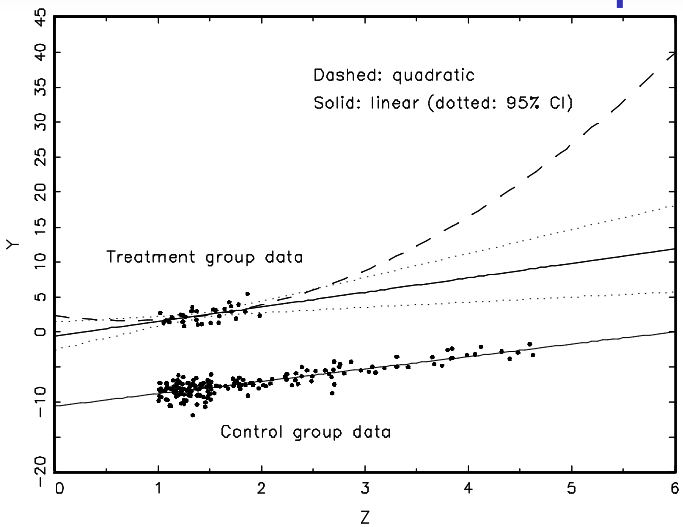
- Must remove data (selecting on X) to avoid extrapolation.
- Options to find "common support" of $p(X|T=1)$ and $P(X|T=0)$
  1. Exact match, so support is defined only at data points

Remove Extrapolation Region, then Match



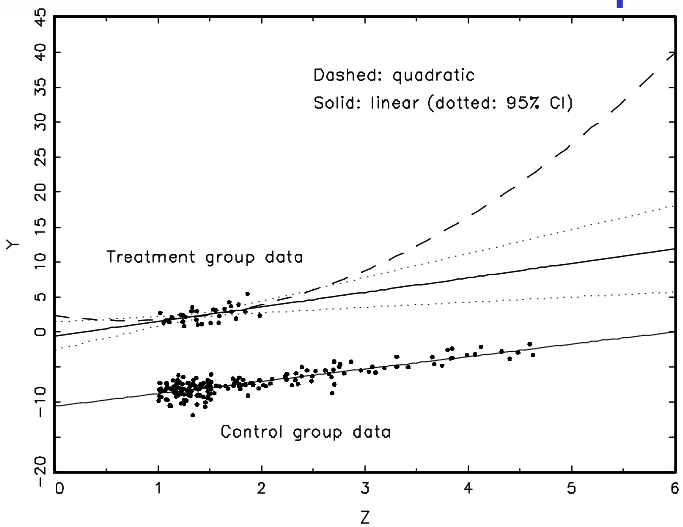
- Must remove data (selecting on X) to avoid extrapolation.
- Options to find "common support" of $p(X|T=1)$ and $P(X|T=0)$
  1. Exact match, so support is defined only at data points
  2. Less but still conservative: convex hull approach

Remove Extrapolation Region, then Match



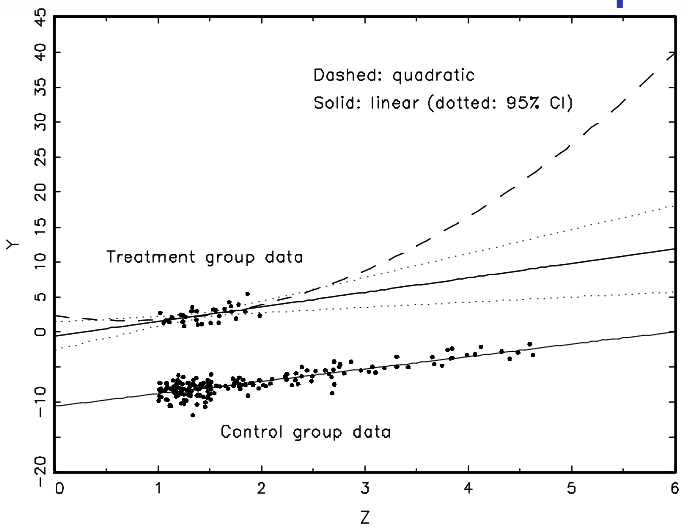
- Must remove data (selecting on X) to avoid extrapolation.
- Options to find "common support" of $p(X|T=1)$ and $P(X|T=0)$
  1. Exact match, so support is defined only at data points
  2. Less but still conservative: convex hull approach
     - let T^* and X^* denote subsets of T and X s.t. $\{1-T^*, X^*\}$ falls within the convex hull of $\{T, X\}$

Remove Extrapolation Region, then Match



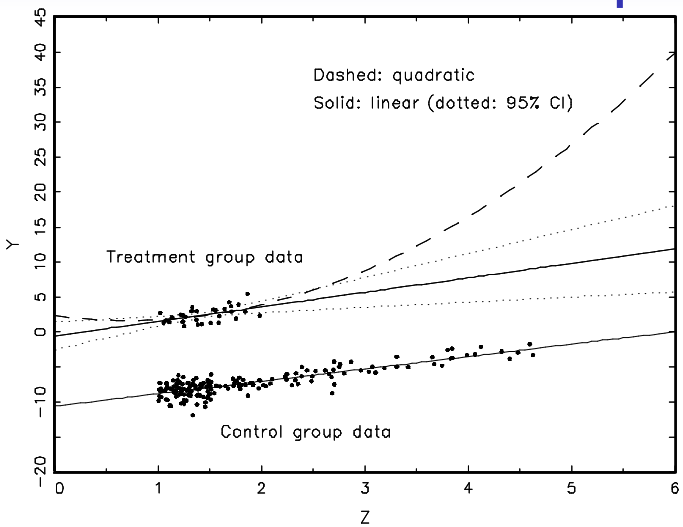
- Must remove data (selecting on X) to avoid extrapolation.
- Options to find "common support" of $p(X|T=1)$ and $P(X|T=0)$
  1. Exact match, so support is defined only at data points
  2. Less but still conservative: convex hull approach
     - let T^* and X^* denote subsets of T and X s.t. $\{1-T^*, X^*\}$ falls within the convex hull of $\{T, X\}$
     - use X^* as estimate of common support (deleting remaining observations)

Remove Extrapolation Region, then Match



- Must remove data (selecting on X) to avoid extrapolation.
- Options to find "common support" of $p(X|T=1)$ and $P(X|T=0)$
  1. Exact match, so support is defined only at data points
  2. Less but still conservative: convex hull approach
     - let T^* and X^* denote subsets of T and X s.t. $\{1-T^*, X^*\}$ falls within the convex hull of $\{T, X\}$
     - use X^* as estimate of common support (deleting remaining observations)
  3. Other approaches, based on distance metrics, pscores, etc.

Remove Extrapolation Region, then Match



- Must remove data (selecting on X) to avoid extrapolation.
- Options to find "common support" of $p(X|T=1)$ and $P(X|T=0)$
  1. Exact match, so support is defined only at data points
  2. Less but still conservative: convex hull approach
     - let T^* and X^* denote subsets of T and X s.t. $\{1-T^*, X^*\}$ falls within the convex hull of $\{T, X\}$
     - use X^* as estimate of common support (deleting remaining observations)
  3. Other approaches, based on distance metrics, pscores, etc.
  4. Easiest: Coarsened Exact Matching, no separate step needed

Matching within the Interpolation Region

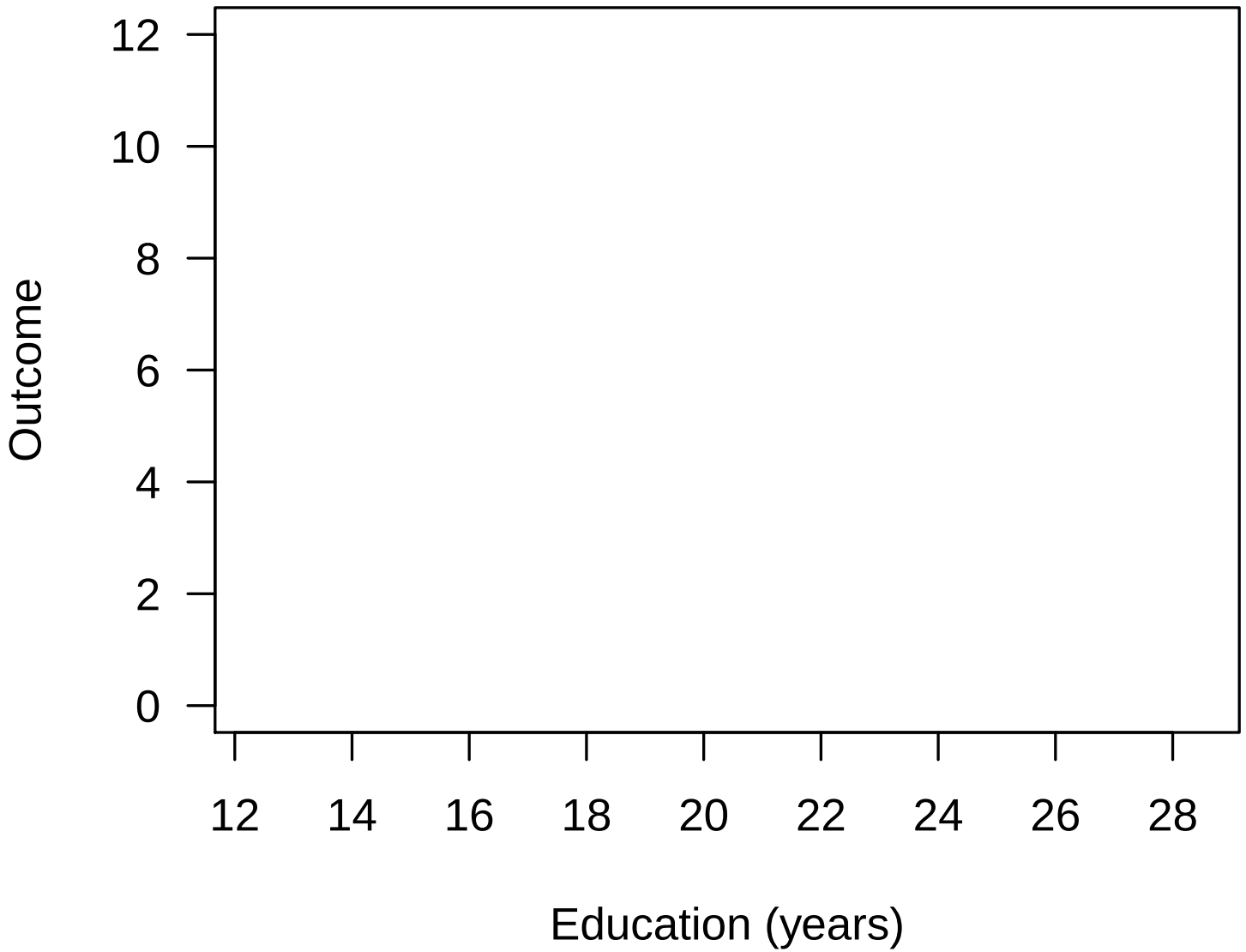
(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)

Matching within the Interpolation Region

(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)

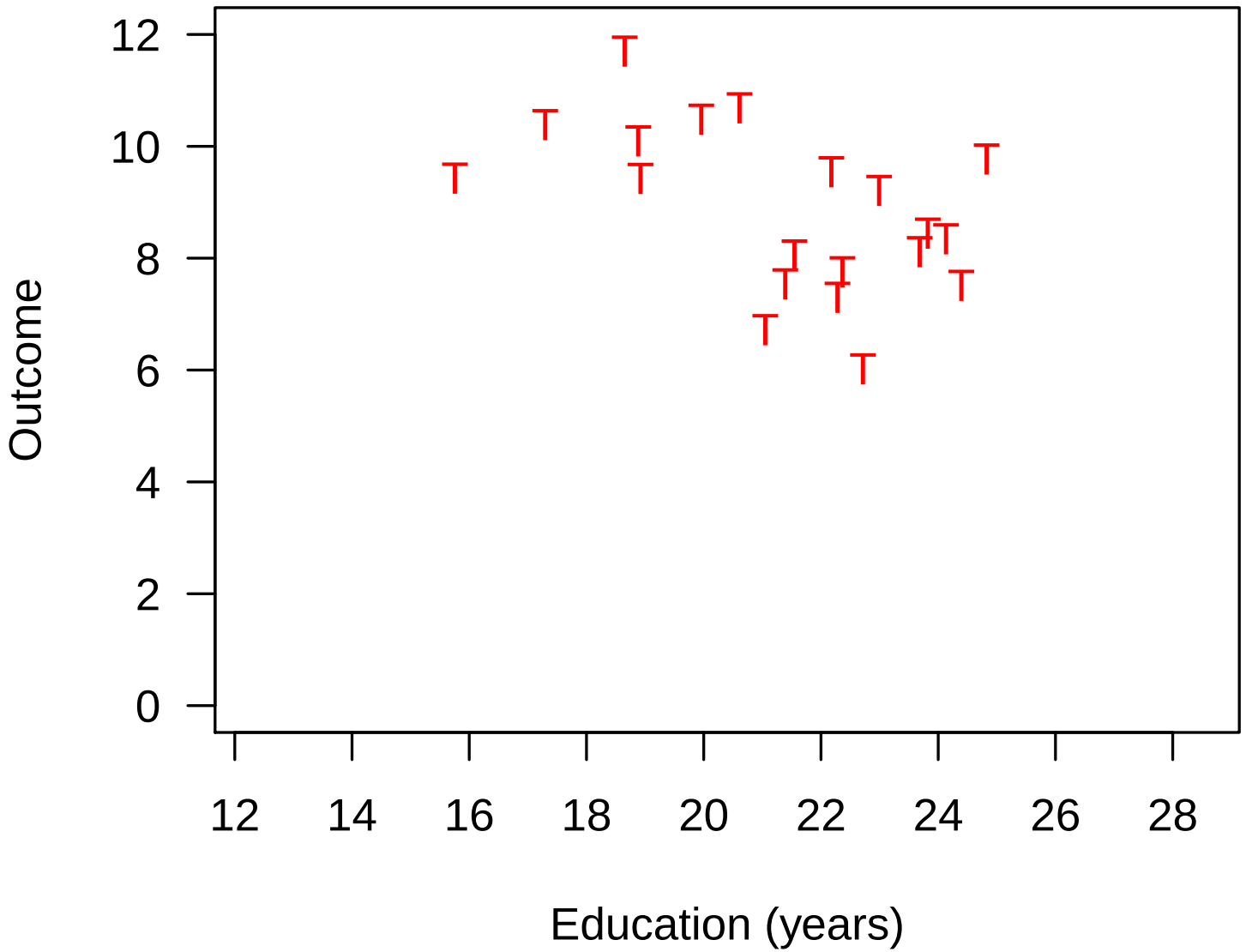
Matching within the Interpolation Region

(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)



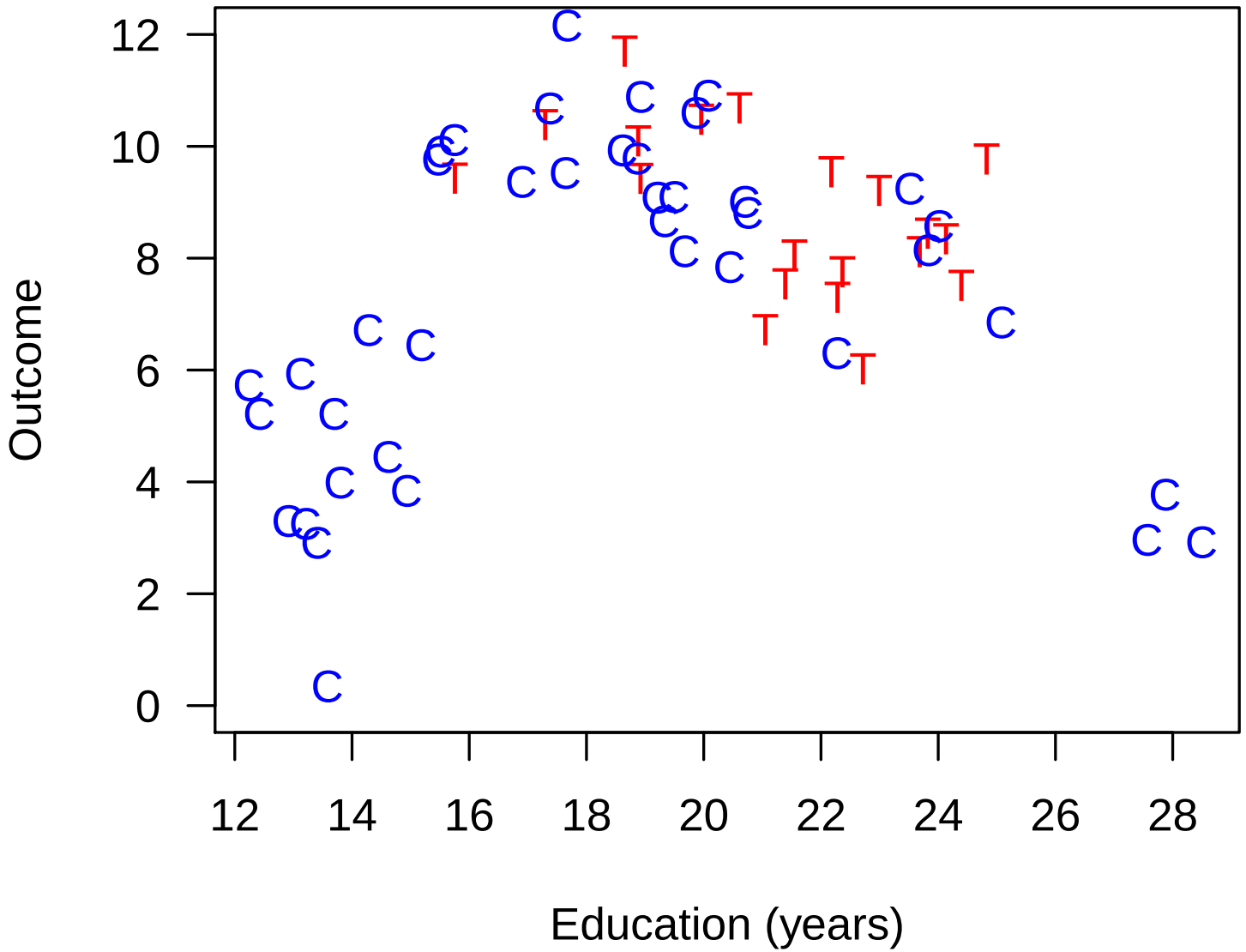
Matching within the Interpolation Region

(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)



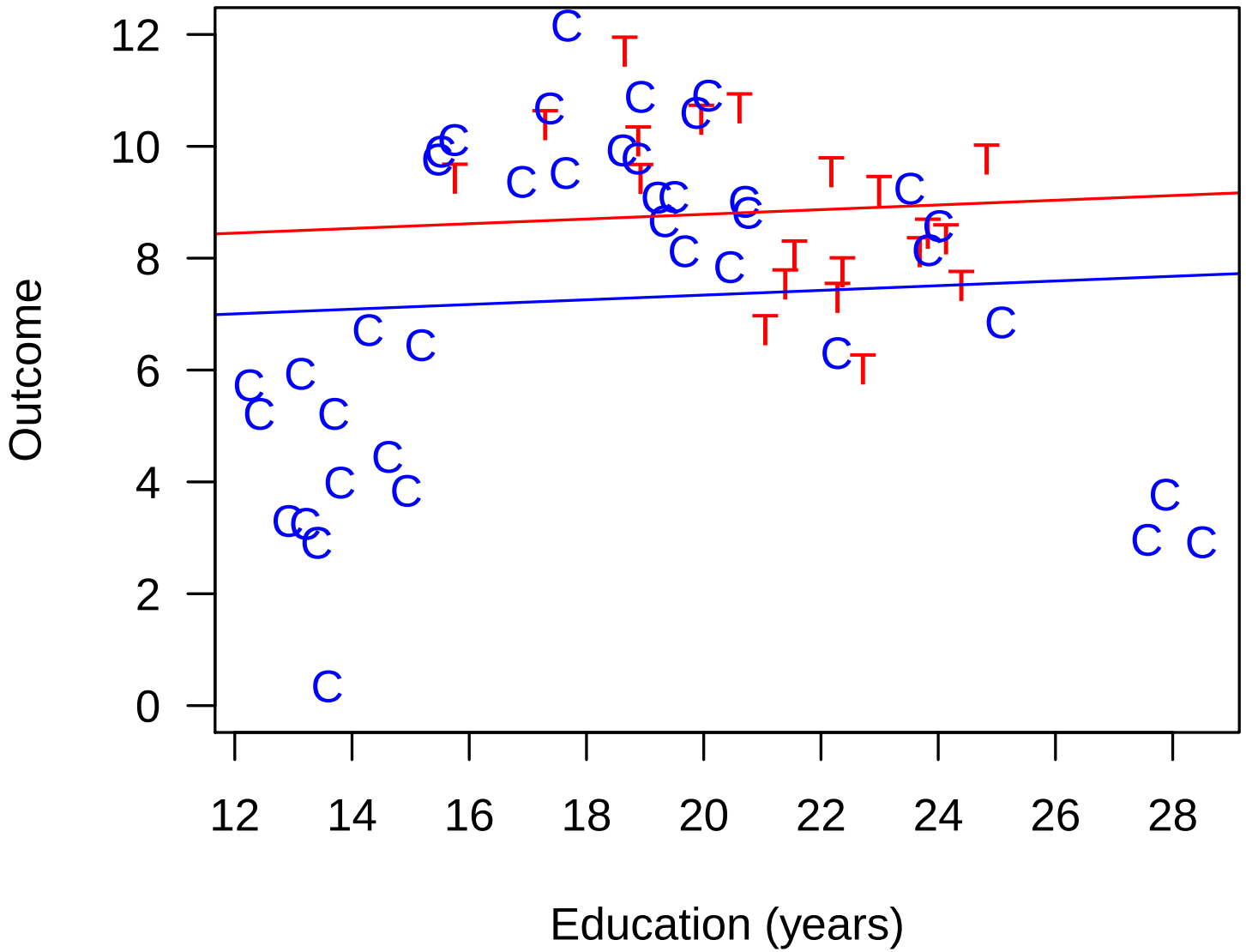
Matching within the Interpolation Region

(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)



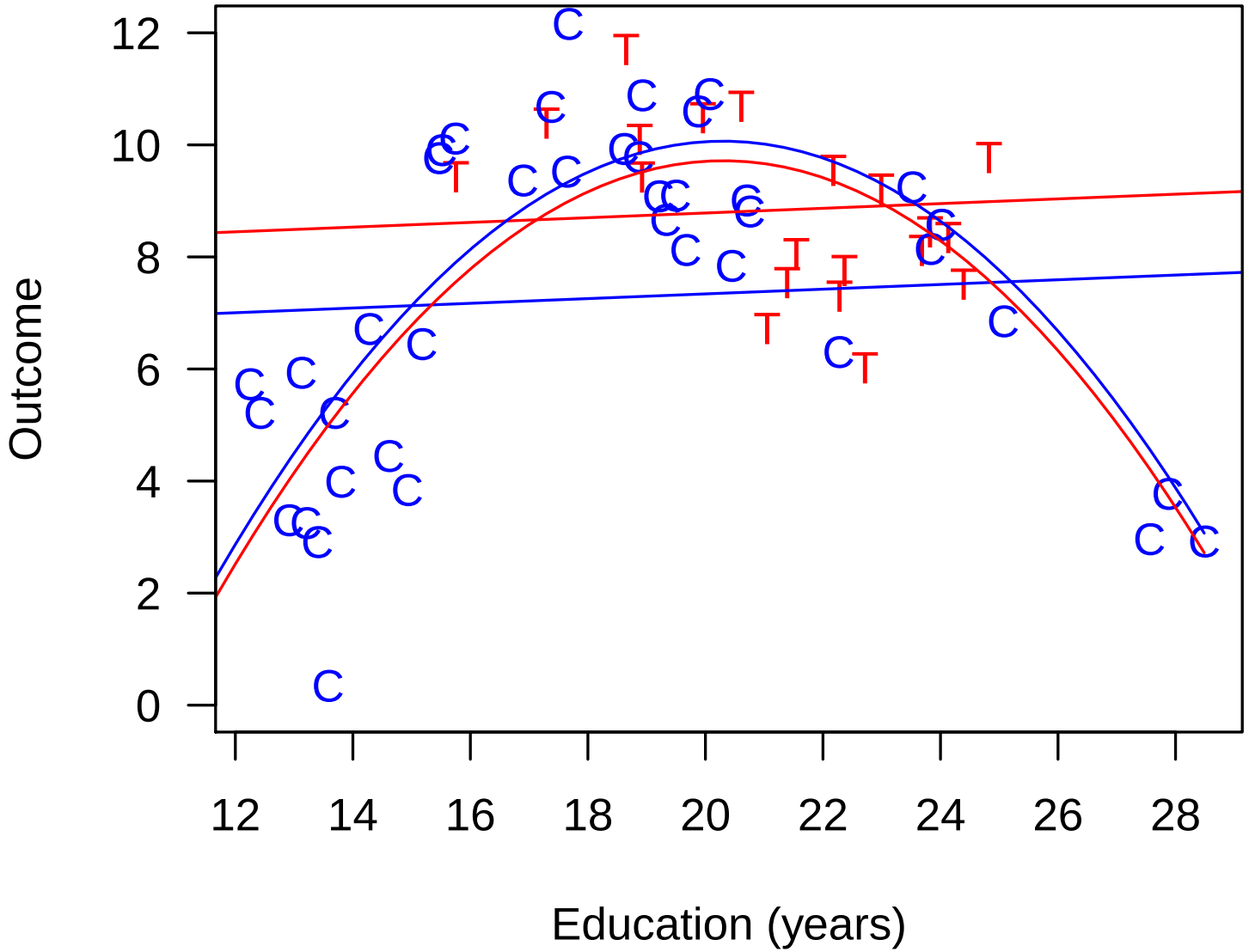
Matching within the Interpolation Region

(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)



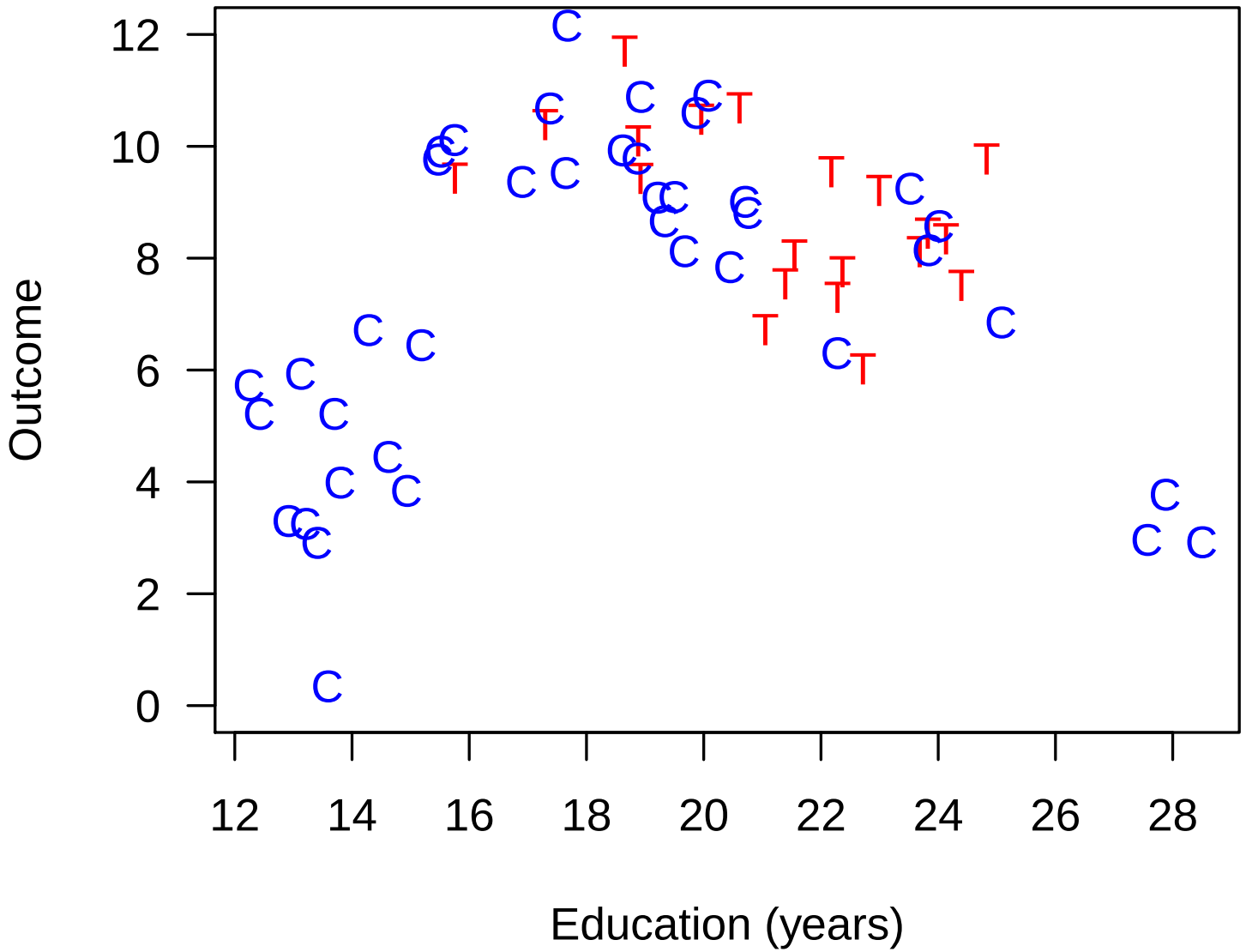
Matching within the Interpolation Region

(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)



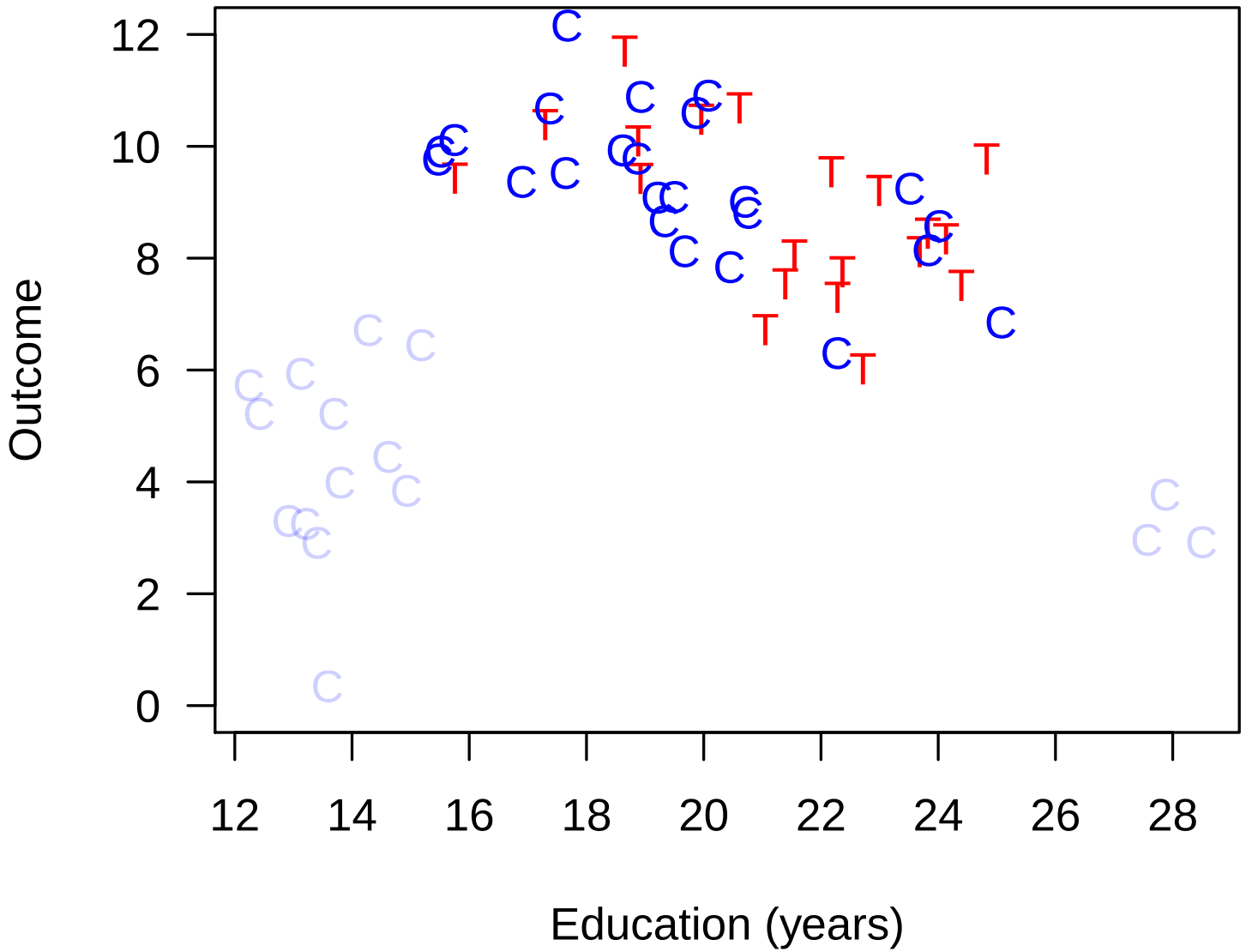
Matching within the Interpolation Region

(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)



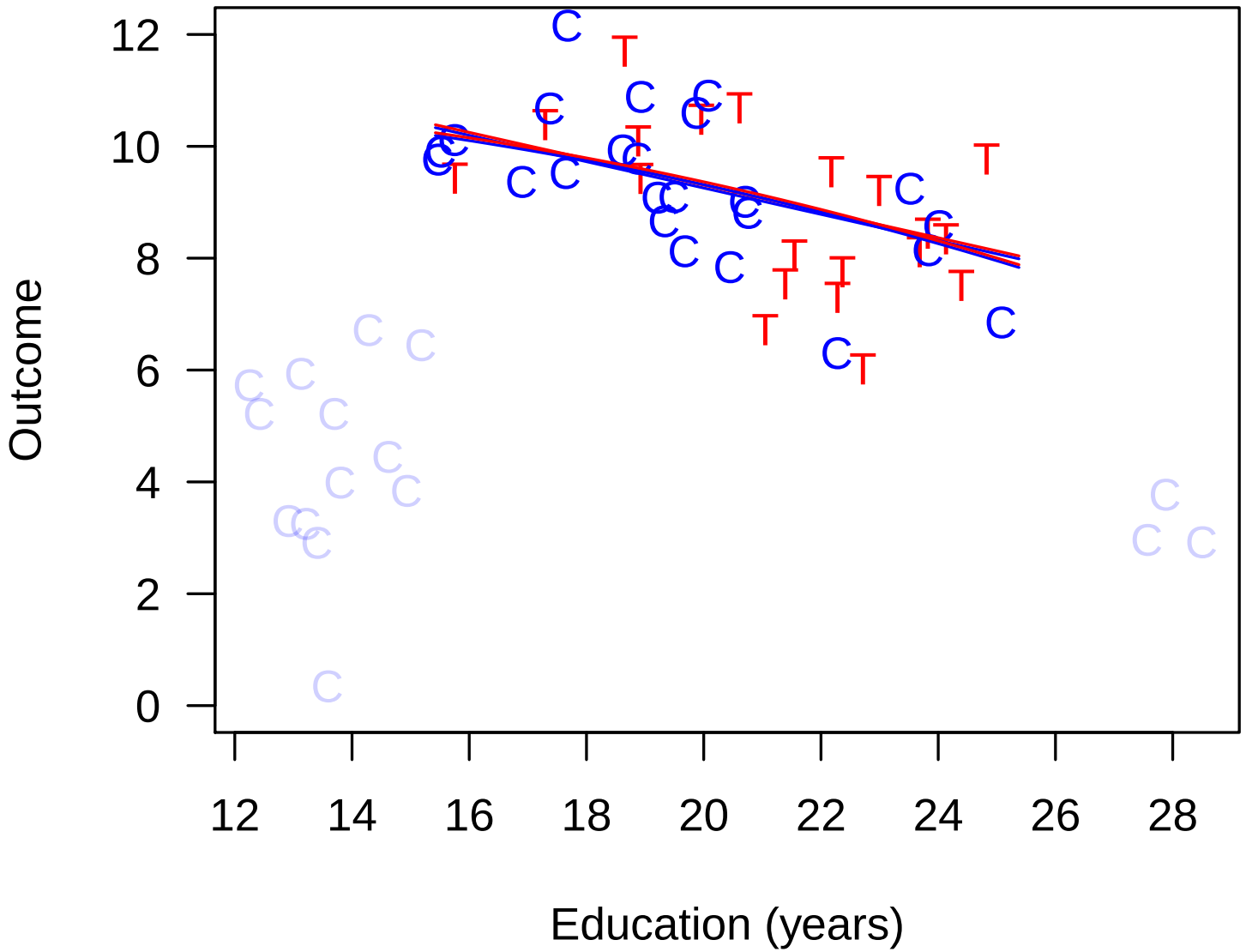
Matching within the Interpolation Region

(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)



Matching within the Interpolation Region

(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)



Matching within the Interpolation Region

(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)

Matching reduces model dependence, bias, and variance

Empirical Illustration: Carpenter, AJPS, 2002

Empirical Illustration: Carpenter, AJPS, 2002

- Hypothesis: Democratic senate majorities slow FDA drug approval time

Empirical Illustration: Carpenter, AJPS, 2002

- Hypothesis: Democratic senate majorities slow FDA drug approval time
- $n = 408$ new drugs (262 approved, 146 pending).

Empirical Illustration: Carpenter, AJPS, 2002

- Hypothesis: Democratic senate majorities slow FDA drug approval time
- $n = 408$ new drugs (262 approved, 146 pending).
- lognormal survival model.

Empirical Illustration: Carpenter, AJPS, 2002

- Hypothesis: Democratic senate majorities slow FDA drug approval time
- $n = 408$ new drugs (262 approved, 146 pending).
- lognormal survival model.
- seven oversight variables (median adjusted ADA scores for House and Senate Committees as well as for House and Senate floors, Democratic Majority in House and Senate, and Democratic Presidency).

Empirical Illustration: Carpenter, AJPS, 2002

- Hypothesis: Democratic senate majorities slow FDA drug approval time
- $n = 408$ new drugs (262 approved, 146 pending).
- lognormal survival model.
- seven oversight variables (median adjusted ADA scores for House and Senate Committees as well as for House and Senate floors, Democratic Majority in House and Senate, and Democratic Presidency).
- 18 control variables (clinical factors, firm characteristics, media variables, etc.)

Evaluating Reduction in Model Dependence

Evaluating Reduction in Model Dependence

- Focus on the causal effect of a Democratic majority in the Senate (identified by Carpenter as not robust).

Evaluating Reduction in Model Dependence

- Focus on the causal effect of a Democratic majority in the Senate (identified by Carpenter as not robust).
- omit post-treatment variables.

Evaluating Reduction in Model Dependence

- Focus on the causal effect of a Democratic majority in the Senate (identified by Carpenter as not robust).
- omit post-treatment variables.
- use one-to-one nearest neighbor propensity score matching.

Evaluating Reduction in Model Dependence

- Focus on the causal effect of a Democratic majority in the Senate (identified by Carpenter as not robust).
- omit post-treatment variables.
- use one-to-one nearest neighbor propensity score matching.
- discard 49 units (2 treated and 17 control units).

Evaluating Reduction in Model Dependence

- Focus on the causal effect of a Democratic majority in the Senate (identified by Carpenter as not robust).
- omit post-treatment variables.
- use one-to-one nearest neighbor propensity score matching.
- discard 49 units (2 treated and 17 control units).
- run 262,143 possible specifications and calculates ATE for each.

Evaluating Reduction in Model Dependence

- Focus on the causal effect of a Democratic majority in the Senate (identified by Carpenter as not robust).
- omit post-treatment variables.
- use one-to-one nearest neighbor propensity score matching.
- discard 49 units (2 treated and 17 control units).
- run 262,143 possible specifications and calculates ATE for each.
- Look at *variability* in ATE estimate across specifications.

Evaluating Reduction in Model Dependence

- Focus on the causal effect of a Democratic majority in the Senate (identified by Carpenter as not robust).
- omit post-treatment variables.
- use one-to-one nearest neighbor propensity score matching.
- discard 49 units (2 treated and 17 control units).
- run 262,143 possible specifications and calculates ATE for each.
- Look at *variability* in ATE estimate across specifications.
- (Normal applications would only use one or a few specifications.)

Reducing Model Dependence

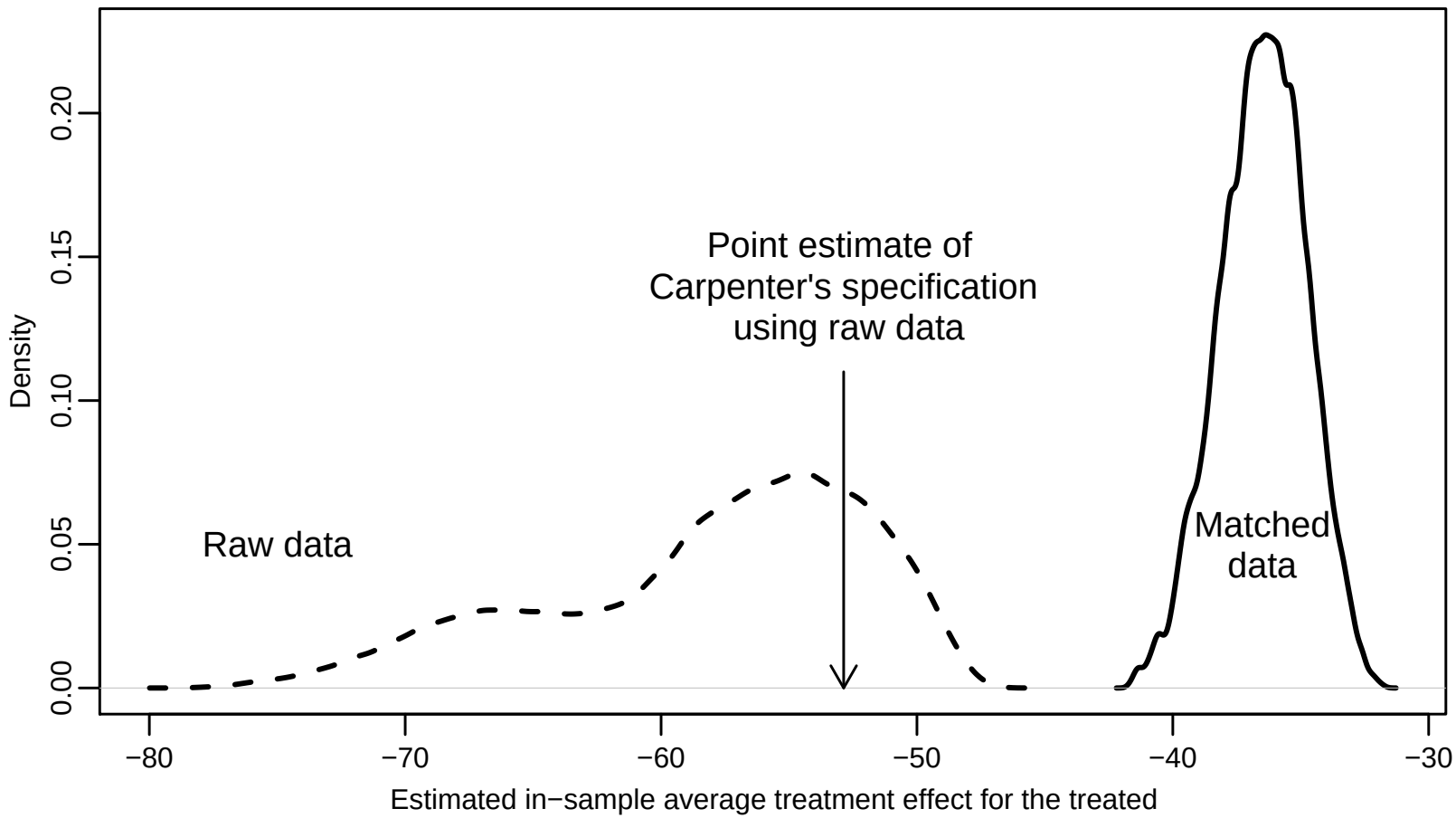


Figure: SATT Histogram: Effect of Democratic Senate majority on FDA drug approval time, across 262,143 specifications.

Another Example: Jeffrey Koch, AJPS, 2002

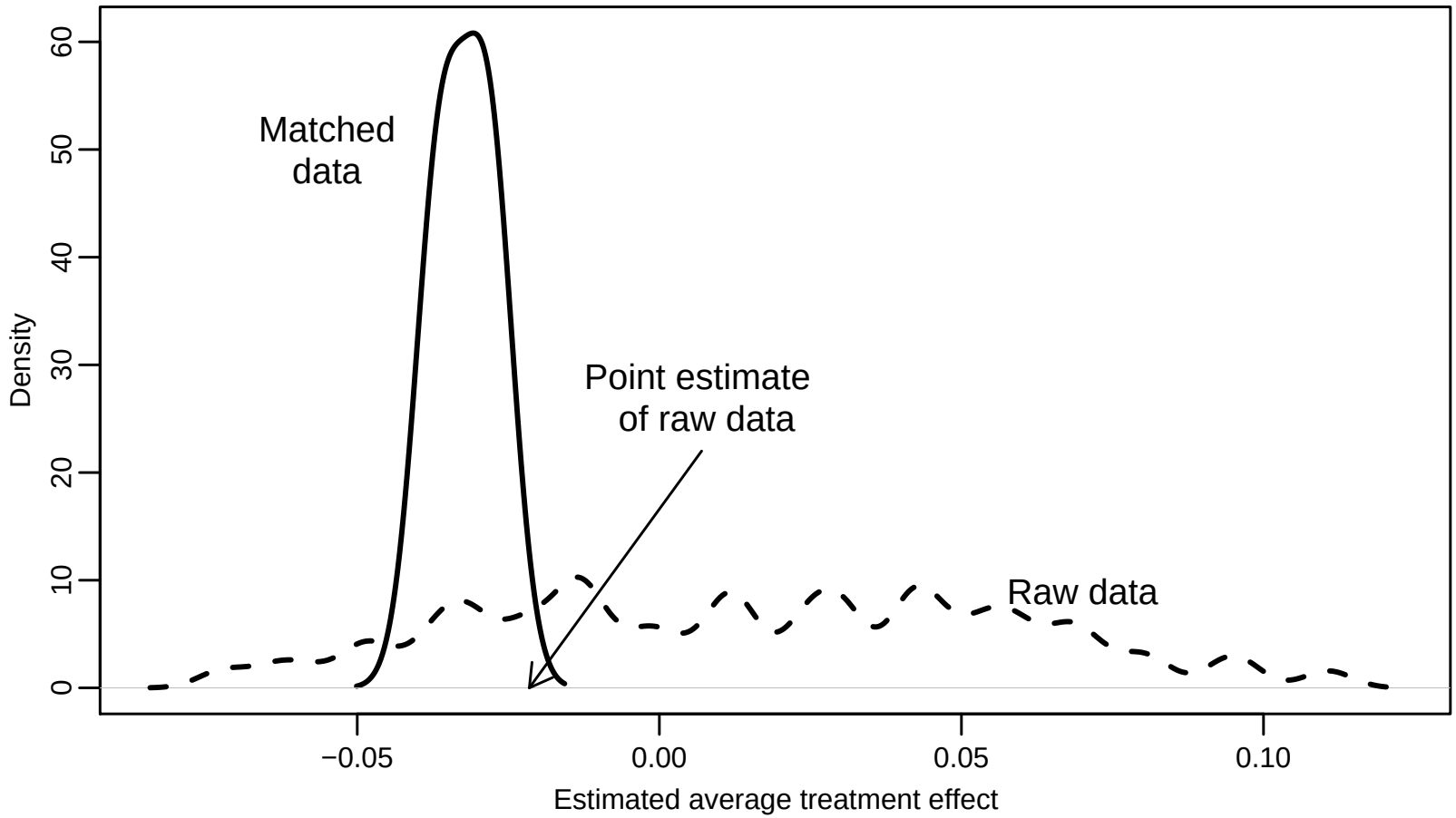


Figure: SATT Histogram: Effect of being a highly visible female Republican candidate across 63 possible specifications with the Koch

How Matching Works

How Matching Works

- Notation:

How Matching Works

- Notation:
  Y_i Dependent variable

How Matching Works

- Notation:
  - Y_i Dependent variable
  - T_i Treatment variable (0/1, or more general)

How Matching Works

- Notation:
  - Y_i Dependent variable
  - T_i Treatment variable (0/1, or more general)
  - X_i Pre-treatment covariates

How Matching Works

- Notation:
  - Y_i Dependent variable
  - T_i Treatment variable (0/1, or more general)
  - X_i Pre-treatment covariates
- Estimation

How Matching Works

- Notation:
  - Y_i Dependent variable
  - T_i Treatment variable (0/1, or more general)
  - X_i Pre-treatment covariates
- Estimation
  - Treatment Effect for treated ($T_i = 1$) observation i:

How Matching Works

- Notation:
  - Y_i Dependent variable
  - T_i Treatment variable (0/1, or more general)
  - X_i Pre-treatment covariates
- Estimation
  - Treatment Effect for treated ($T_i = 1$) observation i:

$$\text{TE}_i = Y_i(T_i = 1) - Y_i(T_i = 0)$$

How Matching Works

- Notation:
  - Y_i Dependent variable
  - T_i Treatment variable (0/1, or more general)
  - X_i Pre-treatment covariates
- Estimation
  - Treatment Effect for treated ($T_i = 1$) observation i:

$$\begin{aligned} \text{TE}_i &= Y_i(T_i = 1) - Y_i(T_i = 0) \\ &= \text{observed} \quad - \textit{unobserved} \end{aligned}$$

How Matching Works

- Notation:
  - Y_i Dependent variable
  - T_i Treatment variable (0/1, or more general)
  - X_i Pre-treatment covariates
- Estimation
  - Treatment Effect for treated ($T_i = 1$) observation i:

$$\begin{aligned} \text{TE}_i &= Y_i(T_i = 1) - Y_i(T_i = 0) \\ &= \text{observed} \quad - \textit{unobserved} \end{aligned}$$

  - Estimate $Y_i(T_i = 0)$ with Y_j from matched ($X_i \approx X_j$) controls $\hat{Y}_i(T_i = 0) = Y_j(T_i = 0)$ (or a model)

How Matching Works

- Notation:
  - Y_i Dependent variable
  - T_i Treatment variable (0/1, or more general)
  - X_i Pre-treatment covariates
- Estimation
  - Treatment Effect for treated ($T_i = 1$) observation i:

$$\begin{aligned} \text{TE}_i &= Y_i(T_i = 1) - Y_i(T_i = 0) \\ &= \text{observed} \quad - \textit{unobserved} \end{aligned}$$

  - Estimate $Y_i(T_i = 0)$ with Y_j from matched ($X_i \approx X_j$) controls $\hat{Y}_i(T_i = 0) = Y_j(T_i = 0)$ (or a model)
  - Prune unmatched units to improve balance (so X is unimportant)

How Matching Works

- Notation:
  - Y_i Dependent variable
  - T_i Treatment variable (0/1, or more general)
  - X_i Pre-treatment covariates
- Estimation
  - Treatment Effect for treated ($T_i = 1$) observation i:

$$\begin{aligned} \text{TE}_i &= Y_i(T_i = 1) - Y_i(T_i = 0) \\ &= \text{observed} \quad - \textit{unobserved} \end{aligned}$$

  - Estimate $Y_i(T_i = 0)$ with Y_j from matched ($X_i \approx X_j$) controls $\hat{Y}_i(T_i = 0) = Y_j(T_i = 0)$ (or a model)
  - Prune unmatched units to improve balance (so X is unimportant)
- Quantities of Interest:

How Matching Works

- Notation:
  - Y_i Dependent variable
  - T_i Treatment variable (0/1, or more general)
  - X_i Pre-treatment covariates
- Estimation
  - Treatment Effect for treated ($T_i = 1$) observation i:

$$\begin{aligned} \text{TE}_i &= Y_i(T_i = 1) - Y_i(T_i = 0) \\ &= \text{observed} \quad - \textit{unobserved} \end{aligned}$$

  - Estimate $Y_i(T_i = 0)$ with Y_j from matched ($X_i \approx X_j$) controls $\hat{Y}_i(T_i = 0) = Y_j(T_i = 0)$ (or a model)
  - Prune unmatched units to improve balance (so X is unimportant)
- Quantities of Interest:
  1. SATT: Sample Average Treatment effect on the Treated:

$$\text{SATT} = \text{mean}_{i \in \{T_i = 1\}} (\text{TE}_i)$$

How Matching Works

- Notation:
  - Y_i Dependent variable
  - T_i Treatment variable (0/1, or more general)
  - X_i Pre-treatment covariates
- Estimation
  - Treatment Effect for treated ($T_i = 1$) observation i:

$$\begin{aligned} \mathrm{TE}_i &= Y_i(T_i = 1) - Y_i(T_i = 0) \\ &= \text{observed} \quad - \textit{unobserved} \end{aligned}$$

  - Estimate $Y_i(T_i = 0)$ with Y_j from matched ($X_i \approx X_j$) controls $\hat{Y}_i(T_i = 0) = Y_j(T_i = 0)$ (or a model)
  - Prune unmatched units to improve balance (so X is unimportant)
- Quantities of Interest:
  1. SATT: Sample Average Treatment effect on the Treated:

$$\mathrm{SATT} = \mathrm{mean}_{i \in \{T_i = 1\}} (\mathrm{TE}_i)$$

  2. FSATT: Feasible Average Treatment effect on the Treated

Method 1: Mahalanobis Distance Matching

Method 1: Mahalanobis Distance Matching

1. Preprocess (Matching)

2. Estimation Difference in means or a model
3. Checking Measure imbalance, tweak, repeat, . . .

Method 1: Mahalanobis Distance Matching

1. Preprocess (Matching)
   - Distance$(X_i, X_j) = \sqrt{(X_i - X_j)' S^{-1} (X_i - X_j)}$

2. Estimation Difference in means or a model
3. Checking Measure imbalance, tweak, repeat, ...

Method 1: Mahalanobis Distance Matching

1. Preprocess (Matching)
   - Distance$(X_i, X_j) = \sqrt{(X_i - X_j)' S^{-1} (X_i - X_j)}$
   - Match each treated unit to the nearest control unit

2. Estimation Difference in means or a model
3. Checking Measure imbalance, tweak, repeat, ...

Method 1: Mahalanobis Distance Matching

1. Preprocess (Matching)
   - Distance$(X_i, X_j) = \sqrt{(X_i - X_j)' S^{-1} (X_i - X_j)}$
   - Match each treated unit to the nearest control unit
   - Control units: not reused; pruned if unused

2. Estimation Difference in means or a model
3. Checking Measure imbalance, tweak, repeat, ...

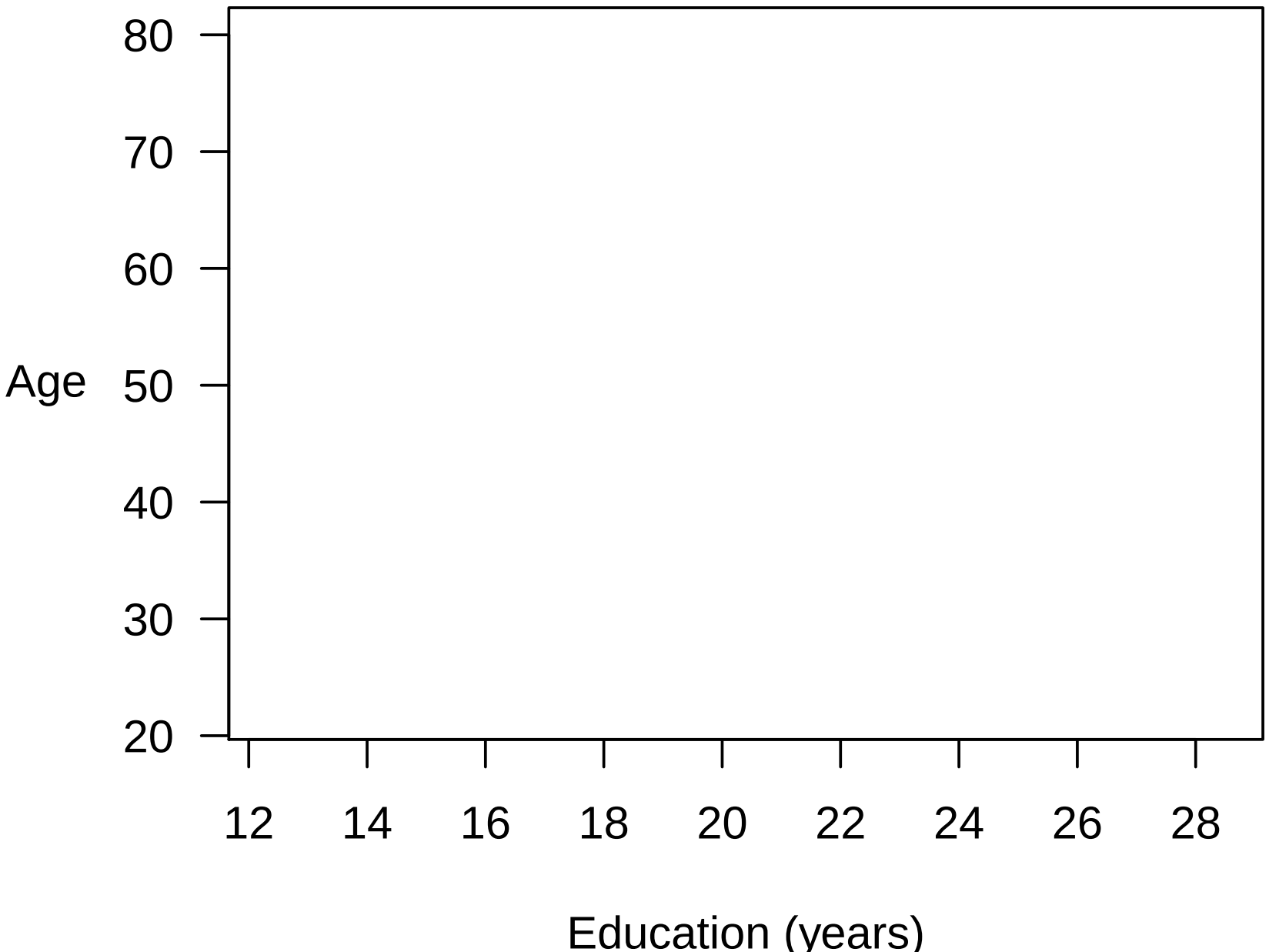
Method 1: Mahalanobis Distance Matching

1. Preprocess (Matching)
   - Distance$(X_i, X_j) = \sqrt{(X_i - X_j)' S^{-1} (X_i - X_j)}$
   - Match each treated unit to the nearest control unit
   - Control units: not reused; pruned if unused
   - Prune matches if Distance$>$*caliper*
2. Estimation Difference in means or a model
3. Checking Measure imbalance, tweak, repeat, ...

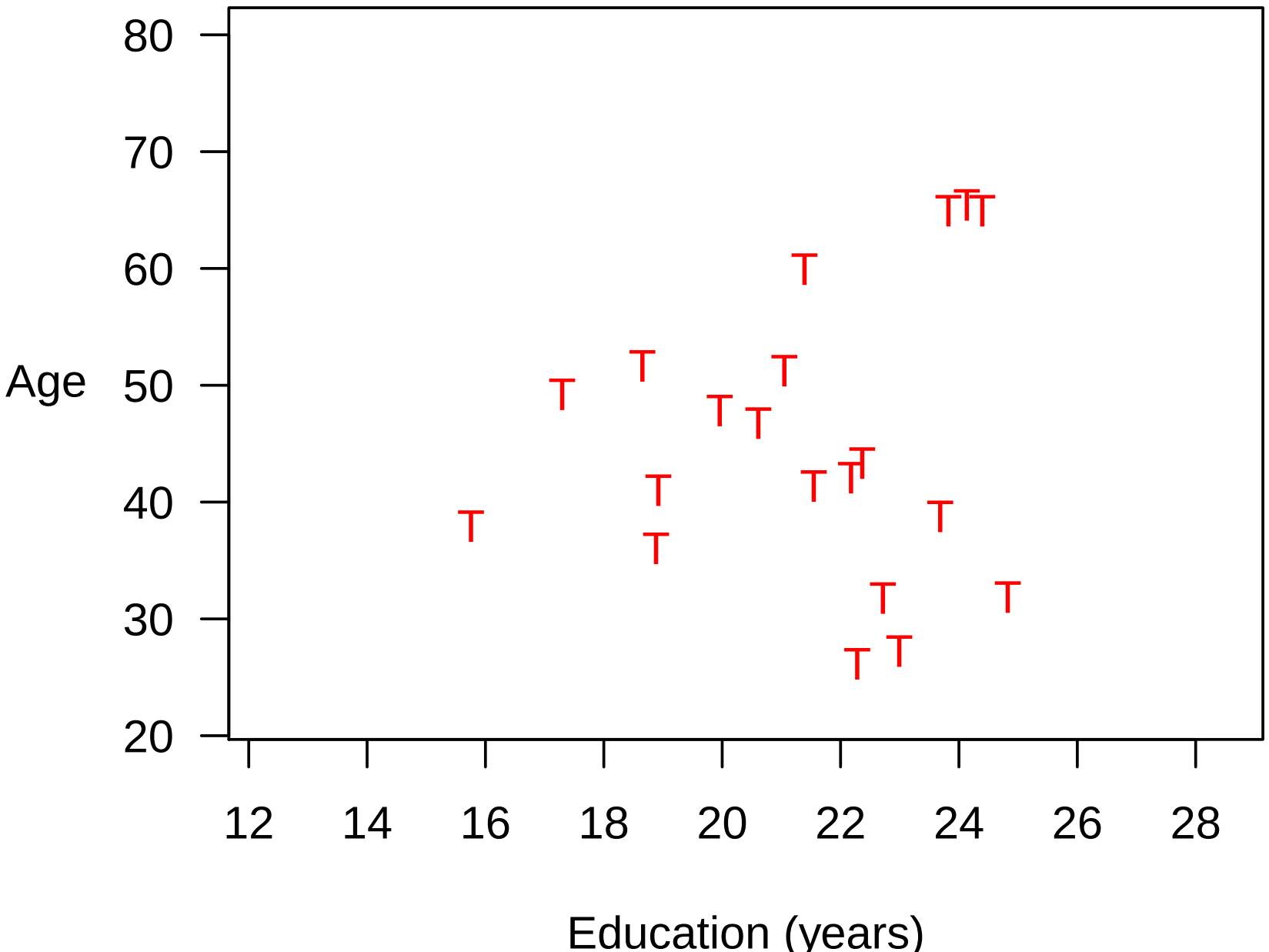
Method 1: Mahalanobis Distance Matching

1. Preprocess (Matching)
   - Distance$(X_i, X_j) = \sqrt{(X_i - X_j)' S^{-1} (X_i - X_j)}$
   - Match each treated unit to the nearest control unit
   - Control units: not reused; pruned if unused
   - Prune matches if Distance$>$*caliper*
2. Estimation Difference in means or a model
3. Checking Measure imbalance, tweak, repeat, ...

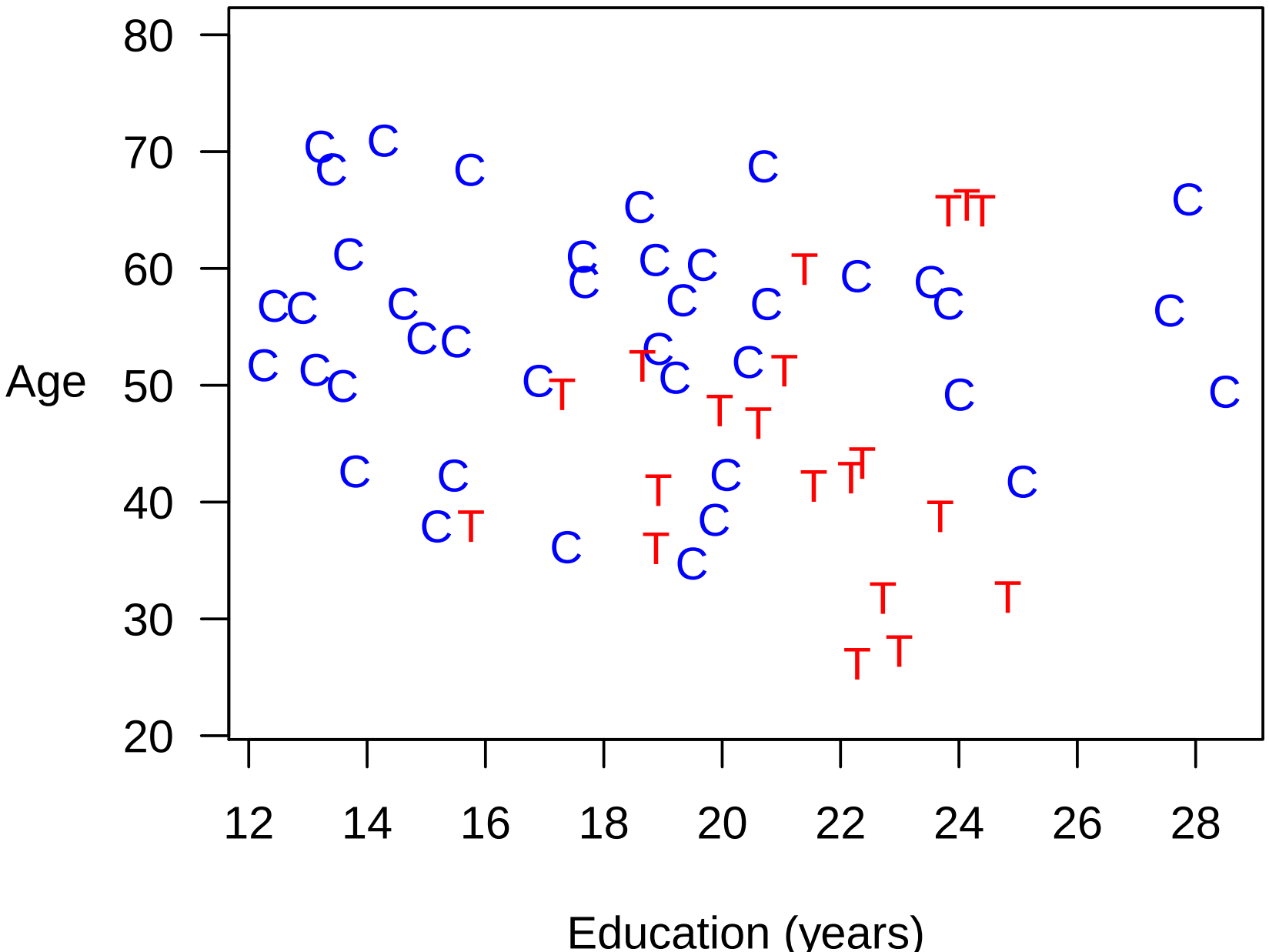
Mahalanobis Distance Matching



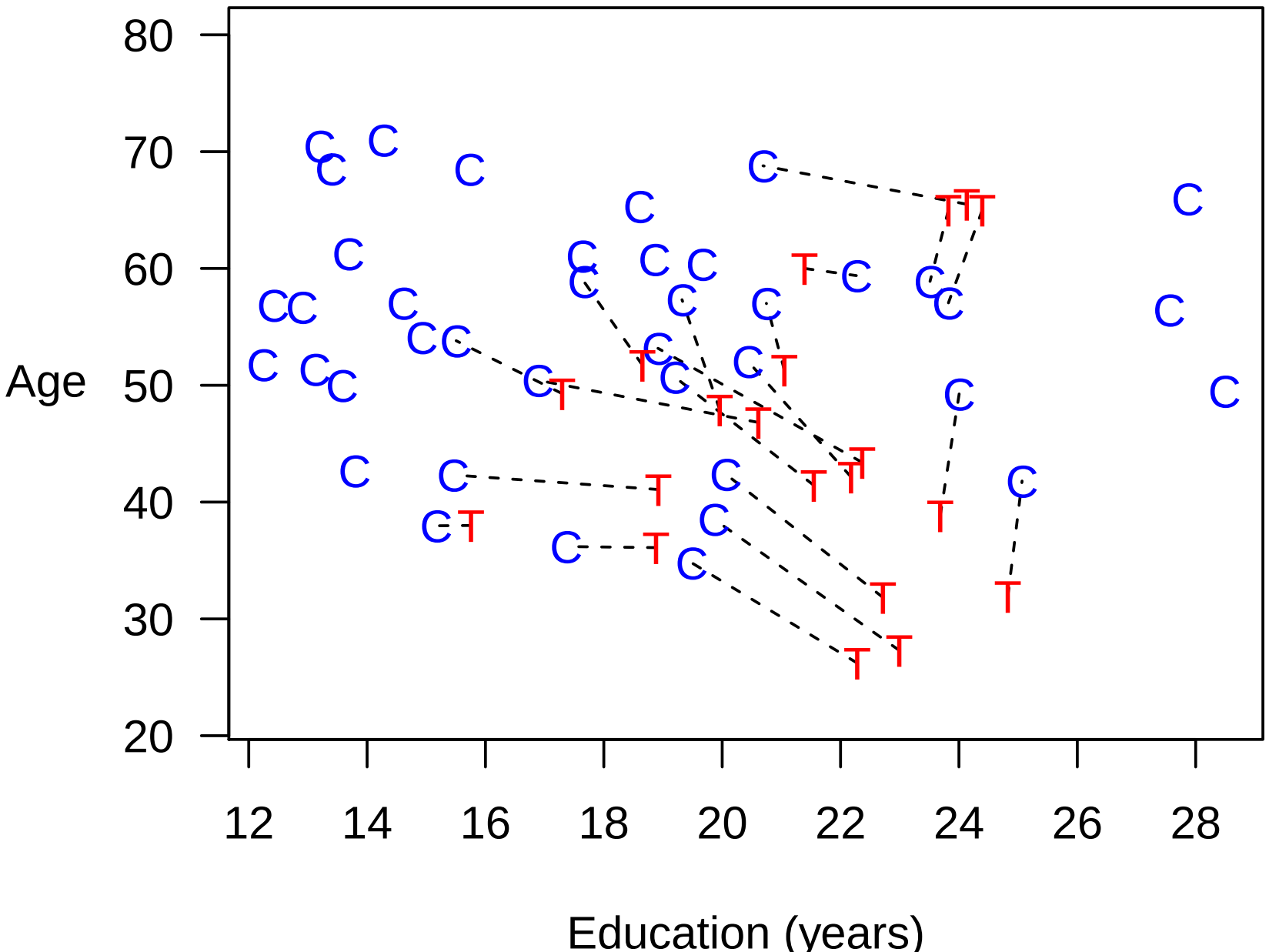
Mahalanobis Distance Matching



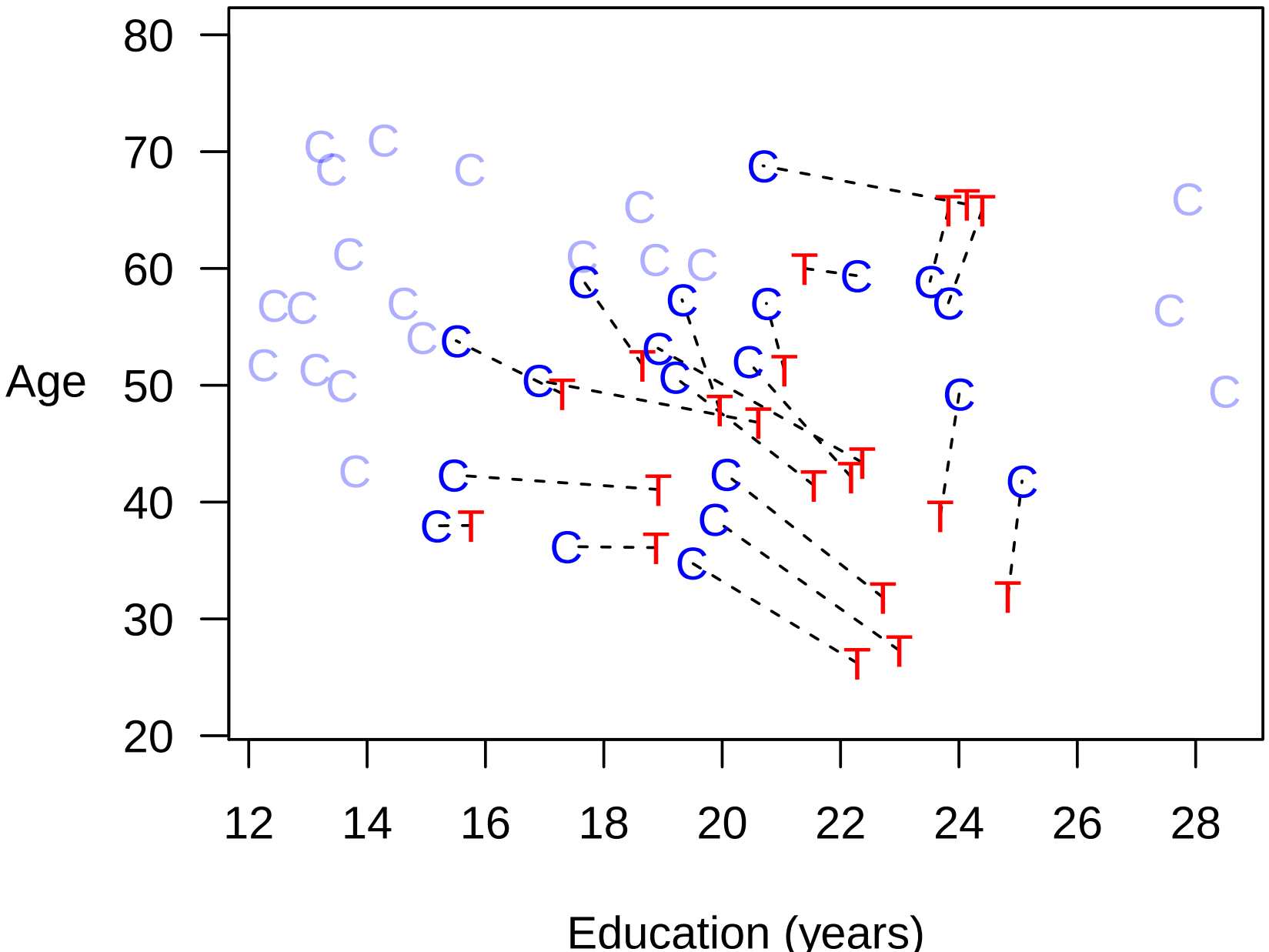
Mahalanobis Distance Matching



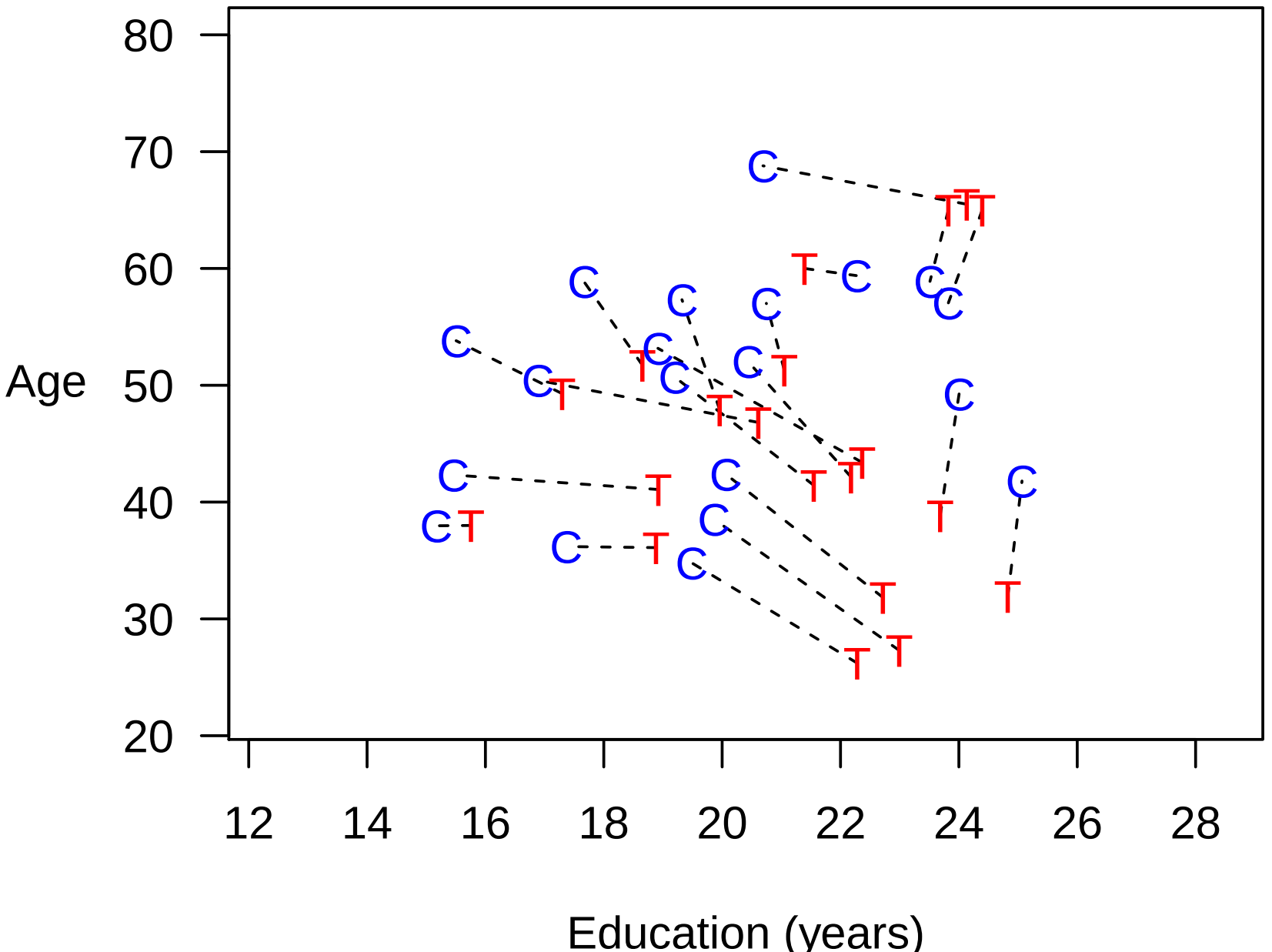
Mahalanobis Distance Matching



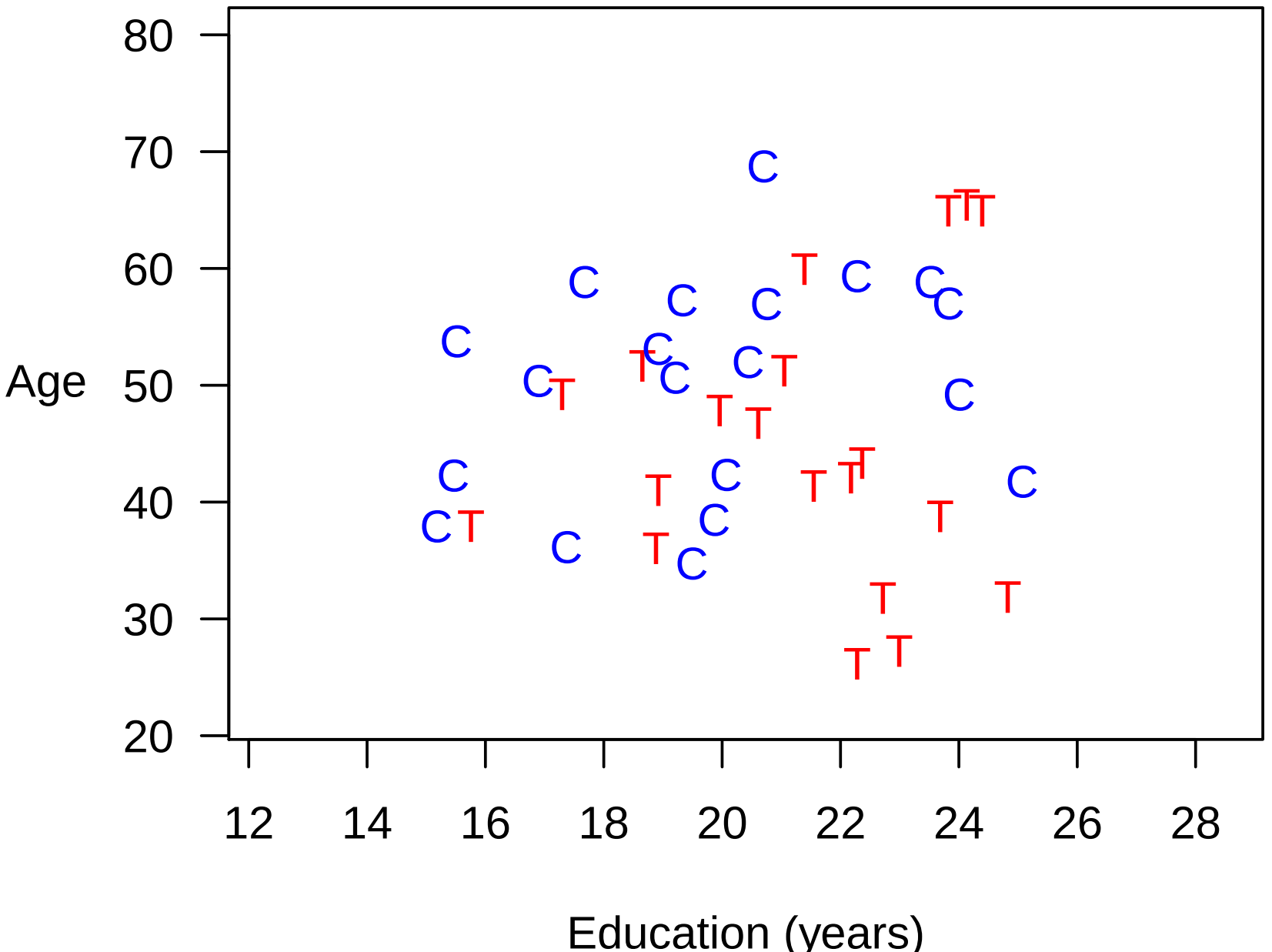
Mahalanobis Distance Matching



Mahalanobis Distance Matching



Mahalanobis Distance Matching



Method 2: Propensity Score Matching

Method 2: Propensity Score Matching

1. Preprocess (Matching)

2. Estimation Difference in means or a model
3. Checking Measure imbalance, tweak, repeat, . . .

Method 2: Propensity Score Matching

1. Preprocess (Matching)
   - Reduce k elements of X to scalar $\pi_i \equiv \Pr(T_i = 1|X) = \frac{1}{1+e^{-X_i\beta}}$
2. Estimation Difference in means or a model
3. Checking Measure imbalance, tweak, repeat, ...

Method 2: Propensity Score Matching

1. Preprocess (Matching)
   - Reduce k elements of X to scalar $\pi_i \equiv \Pr(T_i = 1|X) = \frac{1}{1+e^{-X_i\beta}}$
   - Distance$(X_i, X_j) = |\pi_i - \pi_j|$
2. Estimation Difference in means or a model
3. Checking Measure imbalance, tweak, repeat, ...

Method 2: Propensity Score Matching

1. Preprocess (Matching)
   - Reduce k elements of X to scalar
     $\pi_i \equiv \Pr(T_i = 1|X) = \frac{1}{1+e^{-X_i\beta}}$
   - Distance$(X_i, X_j) = |\pi_i - \pi_j|$
   - Match each treated unit to the nearest control unit

2. Estimation Difference in means or a model
3. Checking Measure imbalance, tweak, repeat, ...

Method 2: Propensity Score Matching

1. Preprocess (Matching)
   - Reduce k elements of X to scalar
     $\pi_i \equiv \Pr(T_i = 1|X) = \frac{1}{1+e^{-X_i\beta}}$
   - Distance$(X_i, X_j) = |\pi_i - \pi_j|$
   - Match each treated unit to the nearest control unit
   - Control units: not reused; pruned if unused

2. Estimation Difference in means or a model
3. Checking Measure imbalance, tweak, repeat, ...

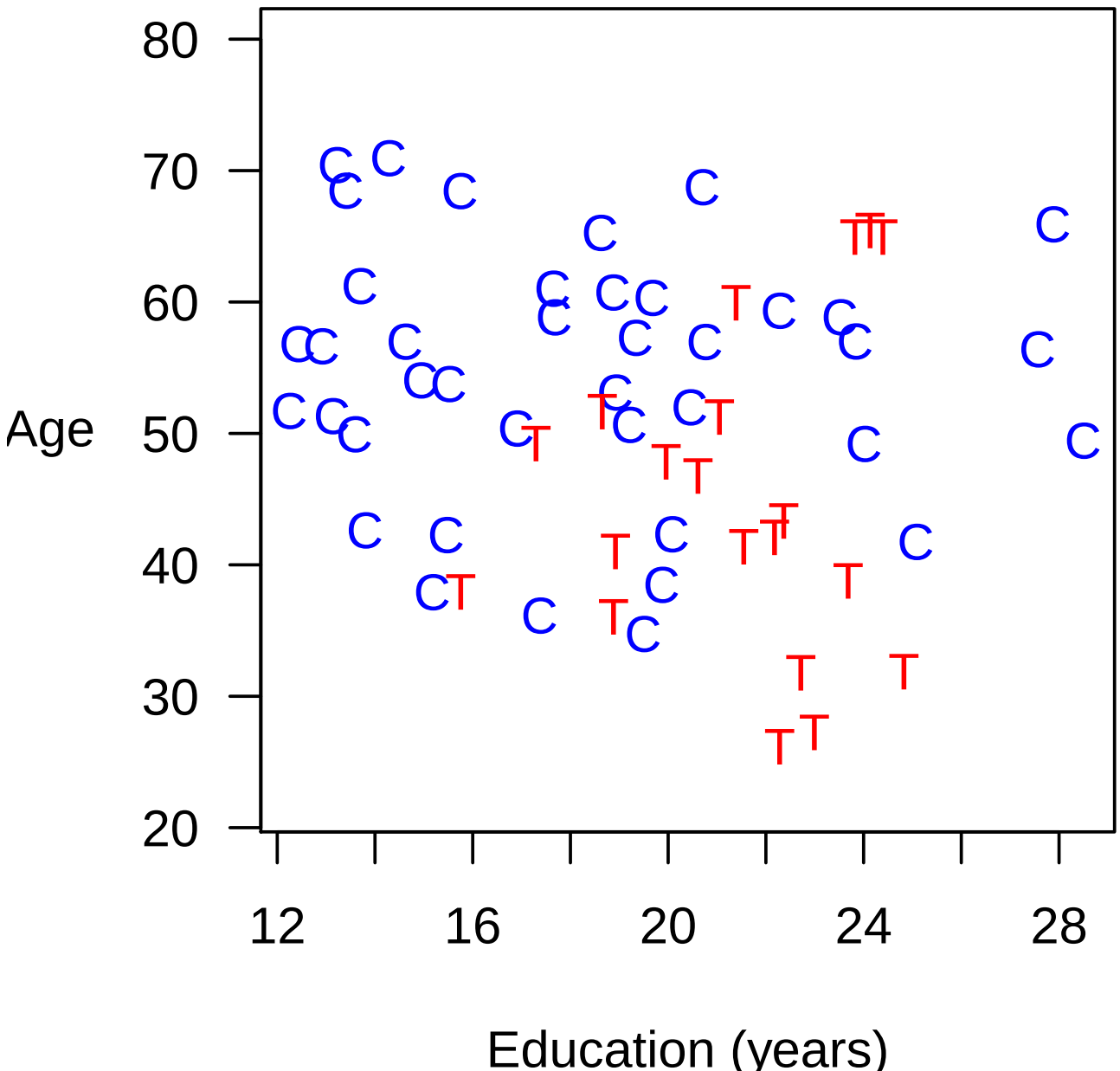
Method 2: Propensity Score Matching

1. Preprocess (Matching)
   - Reduce k elements of X to scalar $\pi_i \equiv \Pr(T_i = 1|X) = \frac{1}{1+e^{-X_i\beta}}$
   - Distance$(X_i, X_j) = |\pi_i - \pi_j|$
   - Match each treated unit to the nearest control unit
   - Control units: not reused; pruned if unused
   - Prune matches if Distance$>$*caliper*
2. Estimation Difference in means or a model
3. Checking Measure imbalance, tweak, repeat, ...

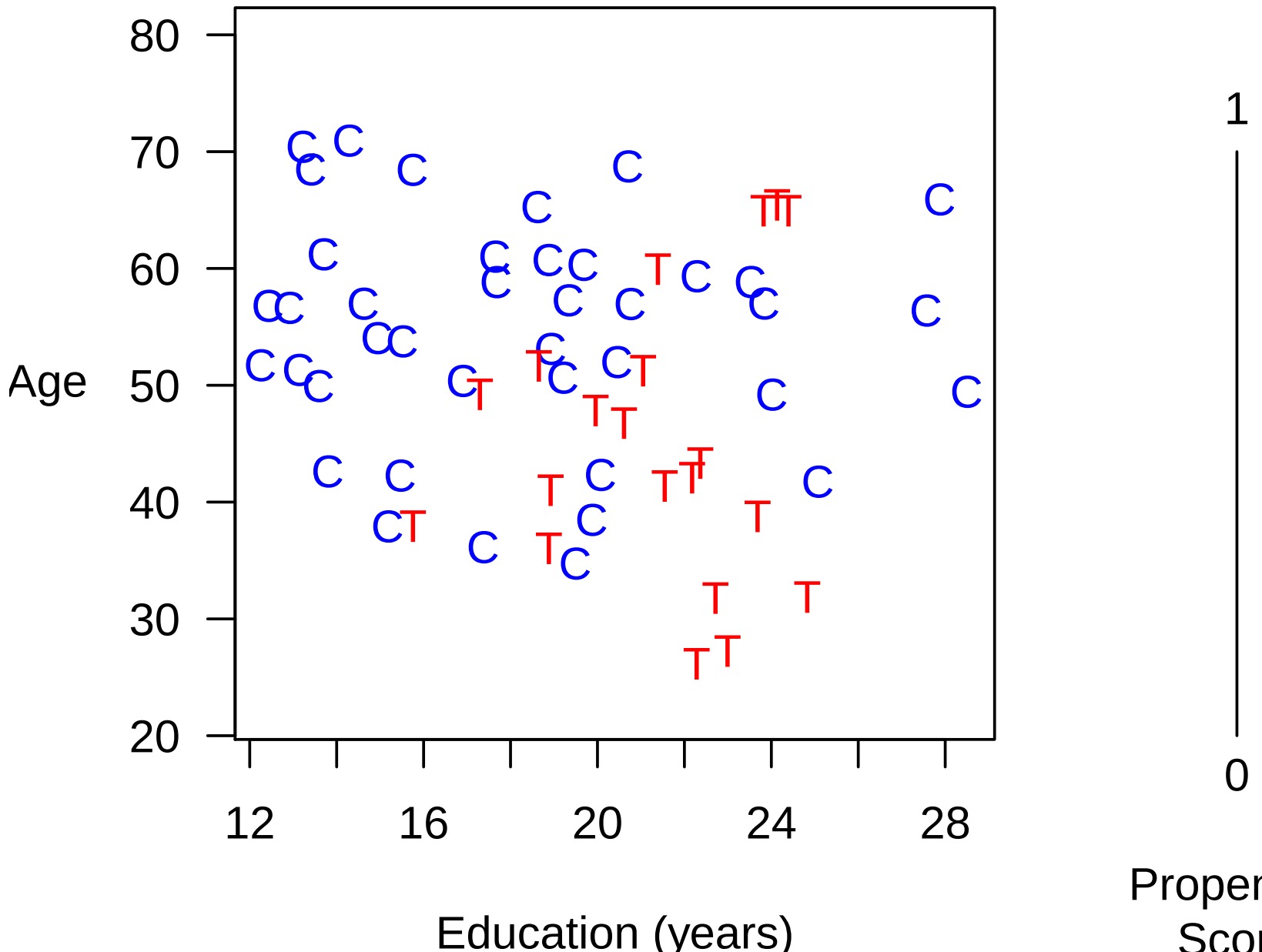
Method 2: Propensity Score Matching

1. Preprocess (Matching)
   - Reduce k elements of X to scalar $\pi_i \equiv \Pr(T_i = 1|X) = \frac{1}{1+e^{-X_i\beta}}$
   - Distance$(X_i, X_j) = |\pi_i - \pi_j|$
   - Match each treated unit to the nearest control unit
   - Control units: not reused; pruned if unused
   - Prune matches if Distance$>$*caliper*
2. Estimation Difference in means or a model
3. Checking Measure imbalance, tweak, repeat, ...

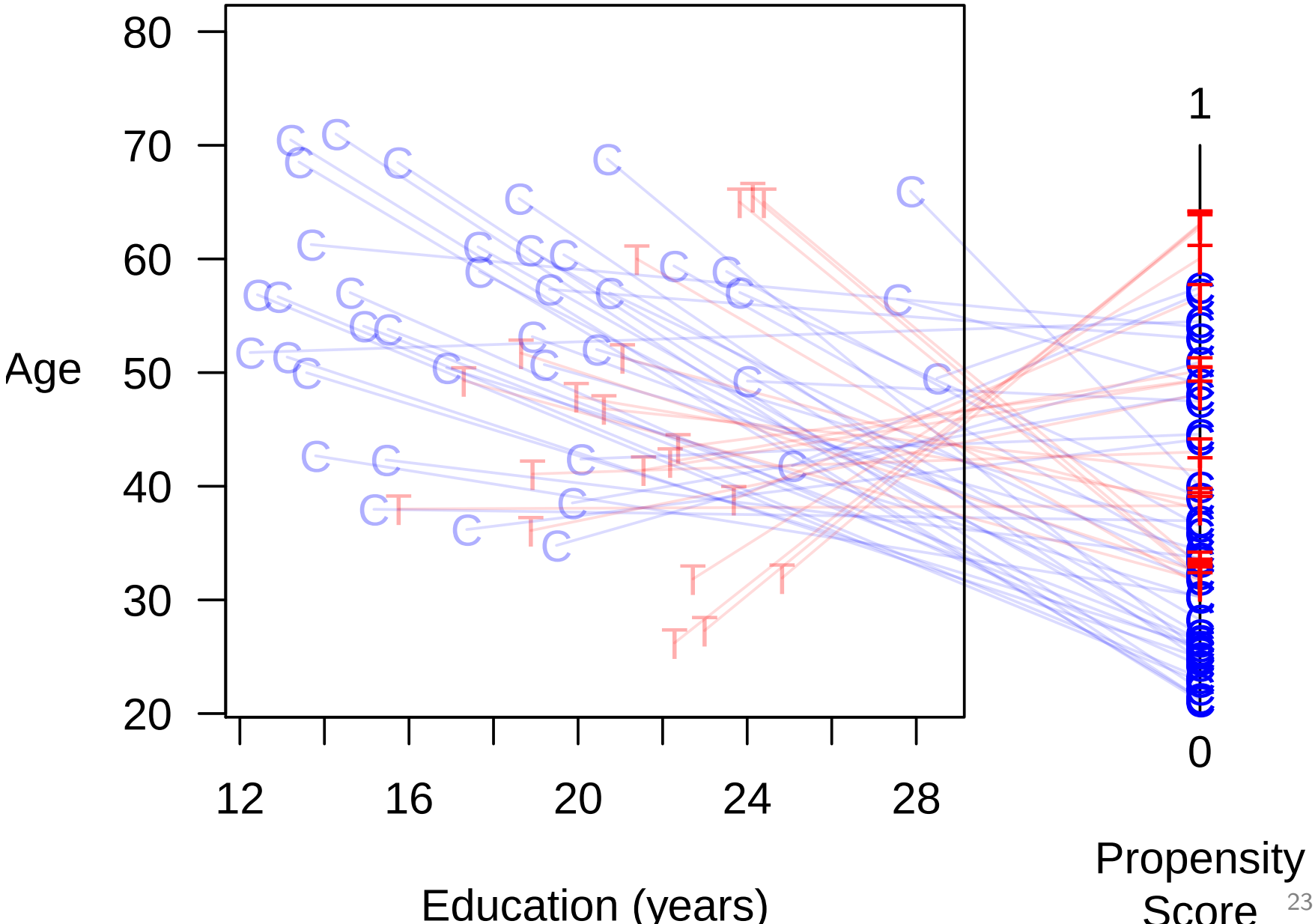
Propensity Score Matching



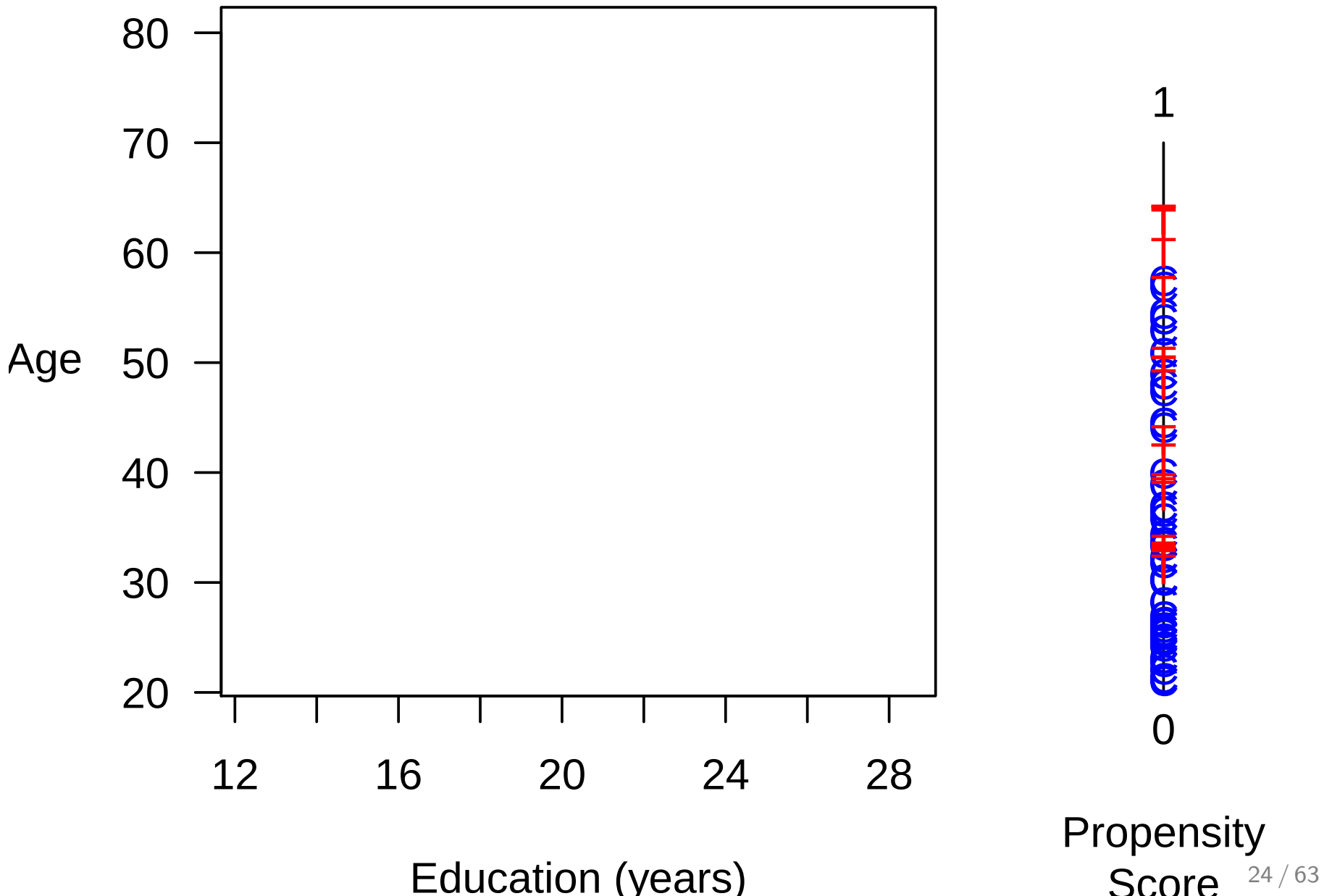
Propensity Score Matching



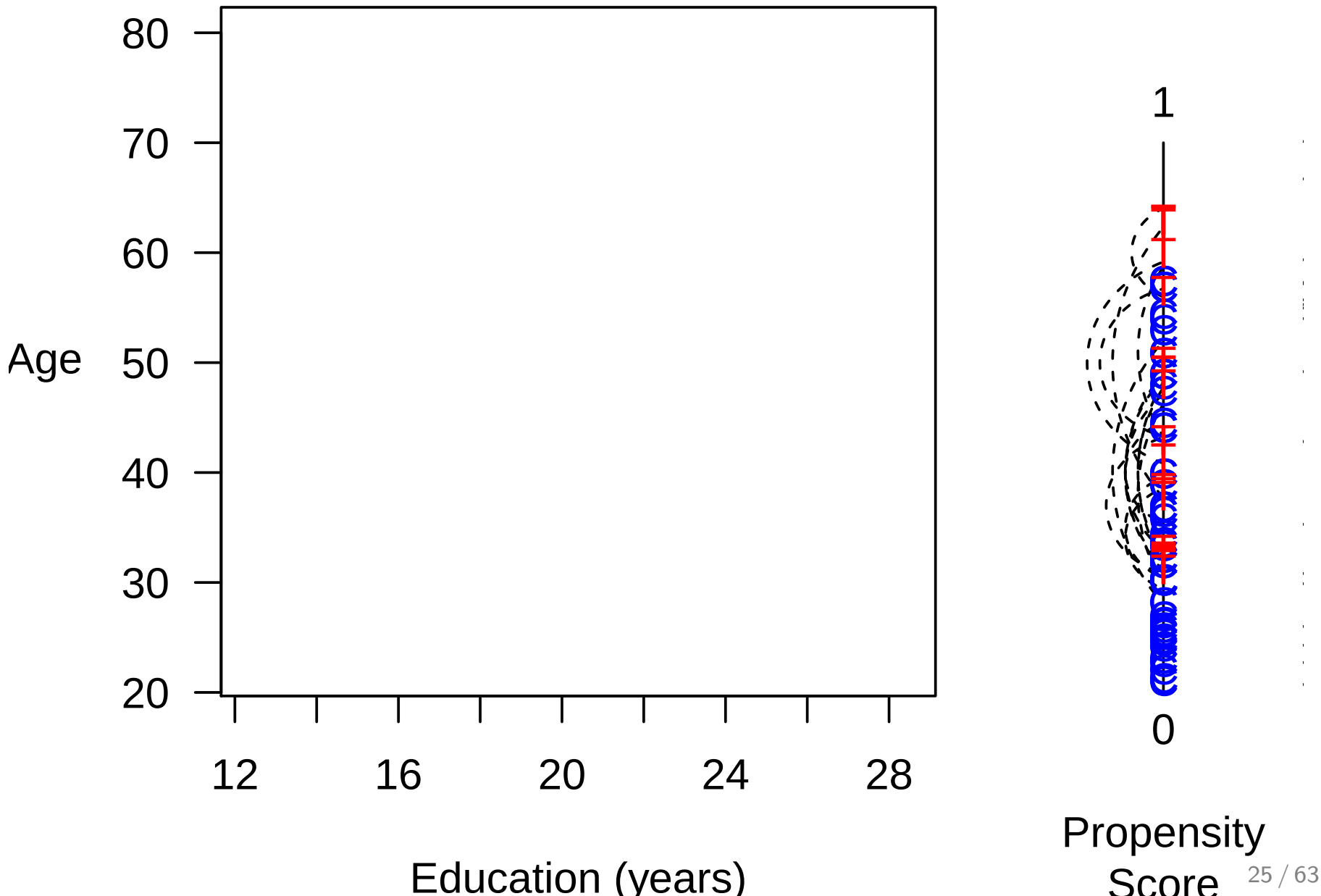
Propensity Score Matching



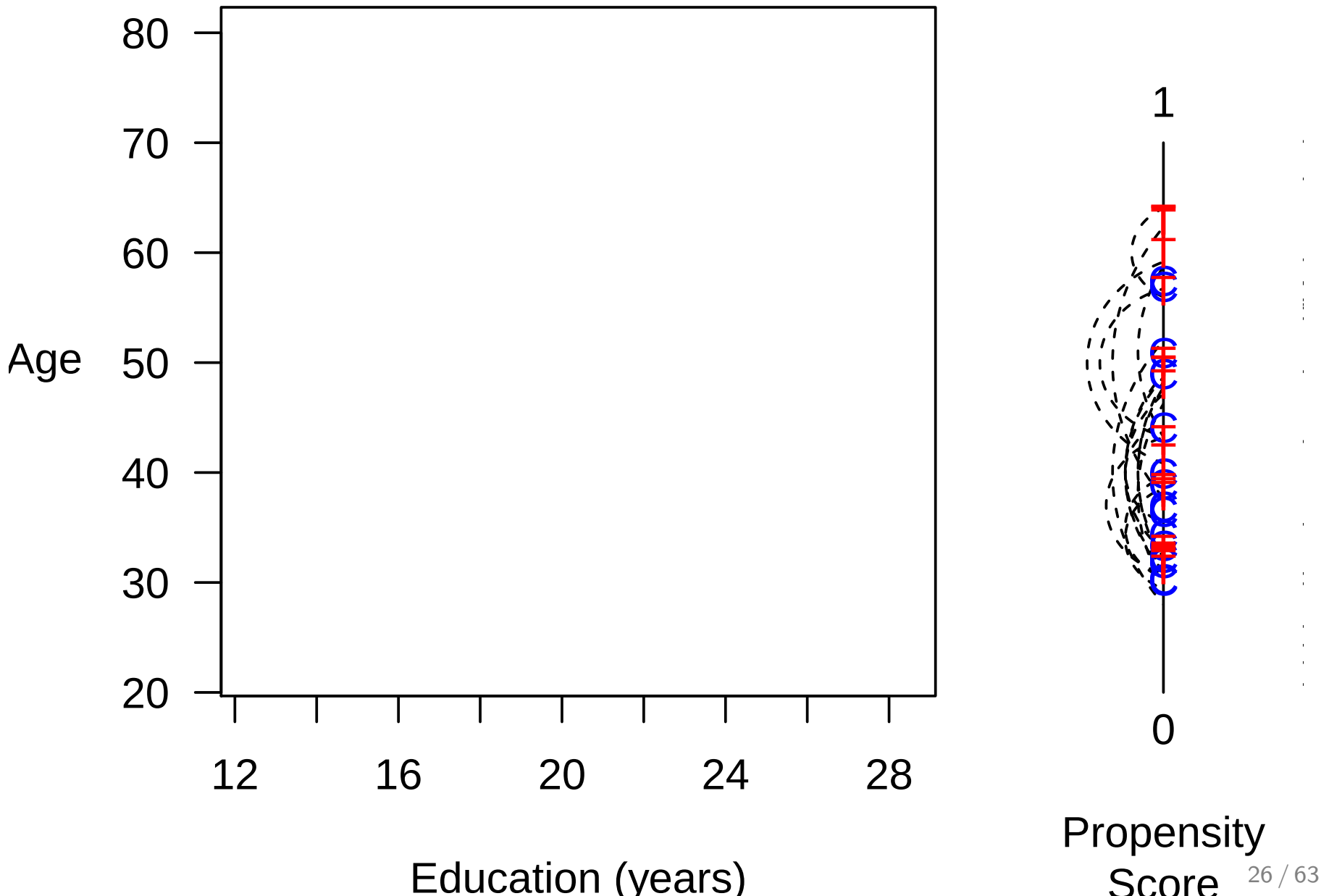
Propensity Score Matching



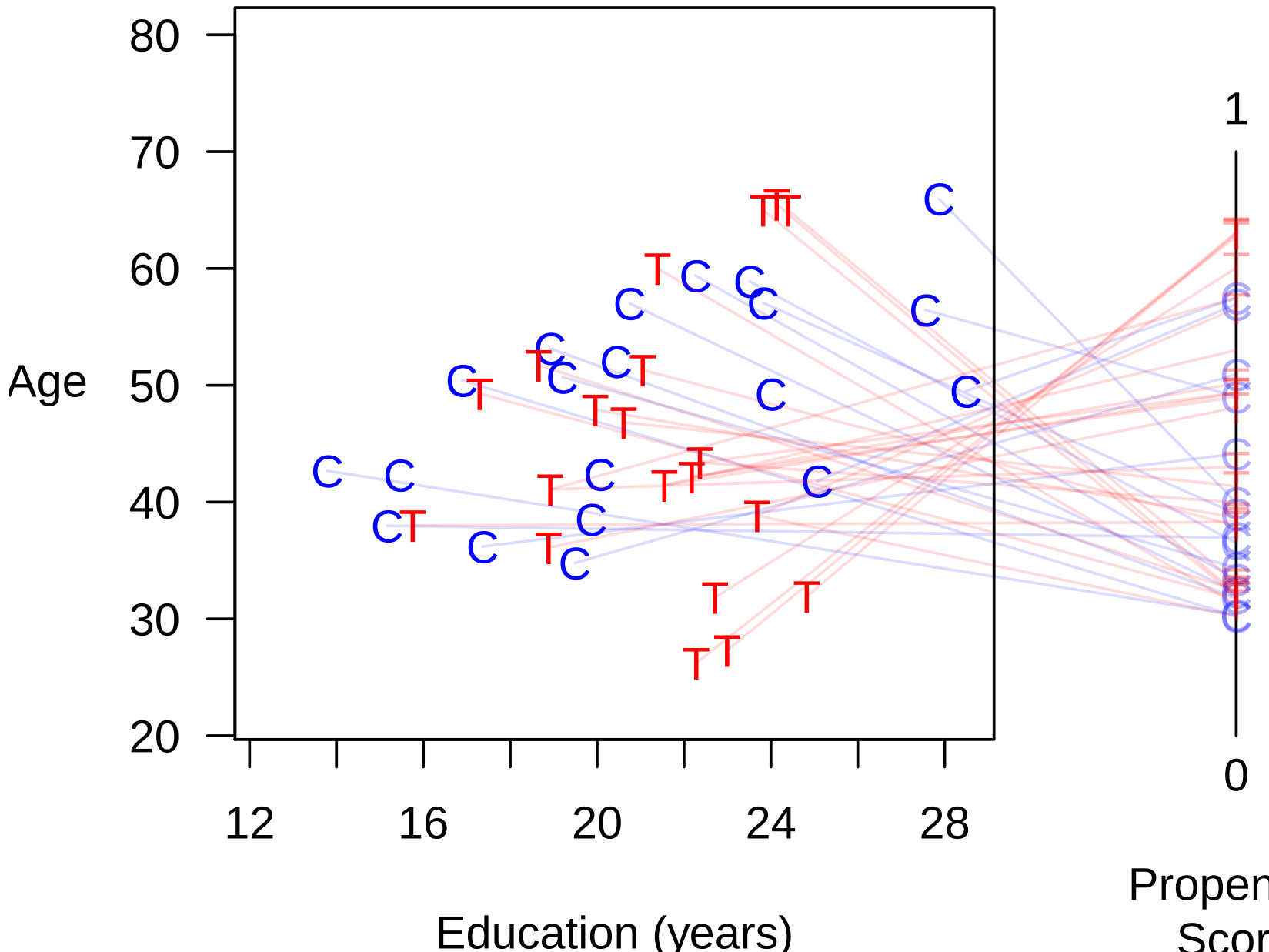
Propensity Score Matching



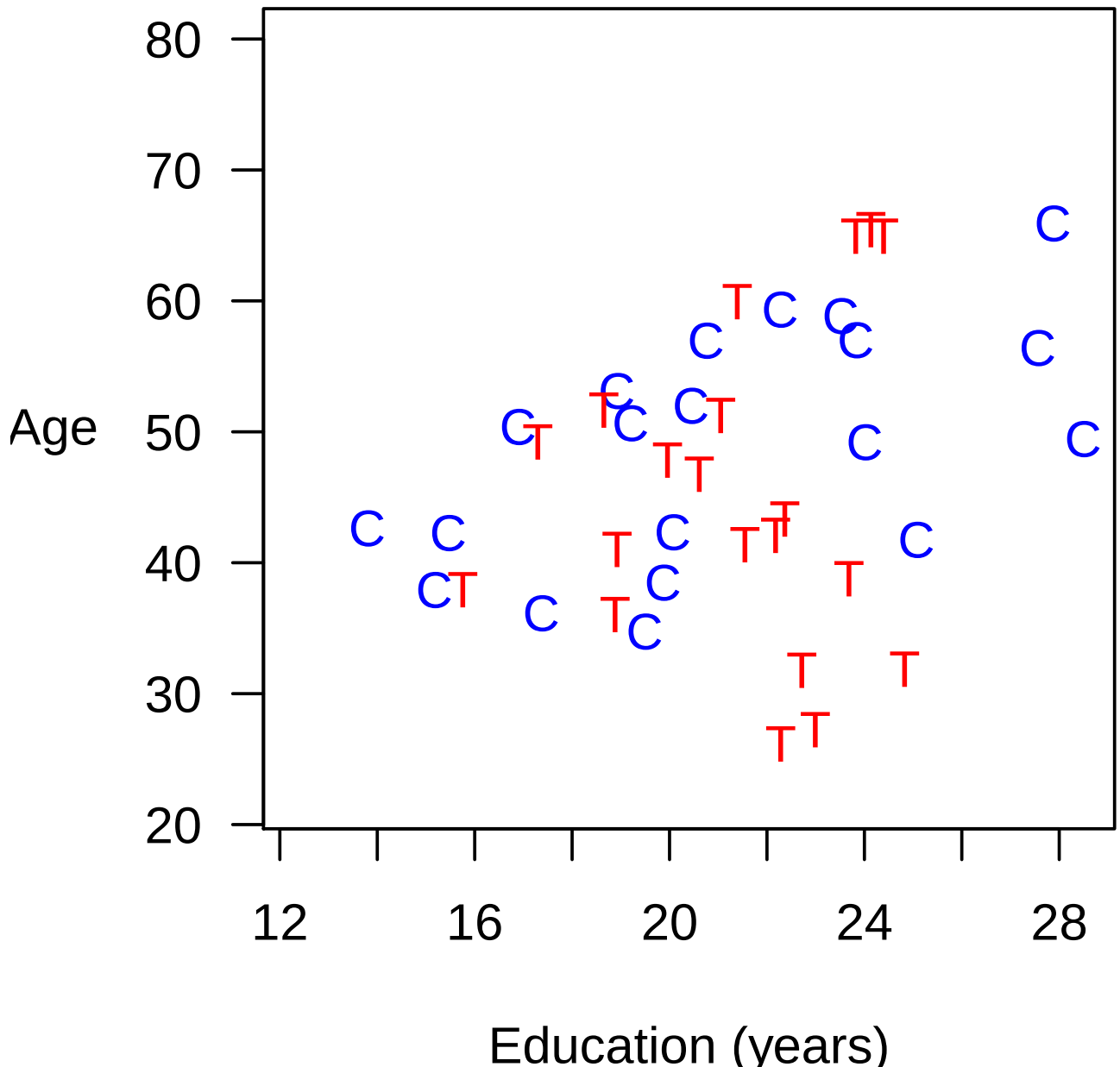
Propensity Score Matching



Propensity Score Matching



Propensity Score Matching



Method 3: Coarsened Exact Matching

Method 3: Coarsened Exact Matching

1. Preprocess (Matching)

2. Estimation Difference in means or a model

3. Checking Determine matched sample size, tweak, repeat, ...

Method 3: Coarsened Exact Matching

1. Preprocess (Matching)
   - Temporarily coarsen X as much as you're willing

2. Estimation Difference in means or a model

3. Checking Determine matched sample size, tweak, repeat, ...

Method 3: Coarsened Exact Matching

1. Preprocess (Matching)
   - Temporarily coarsen X as much as you're willing
     - e.g., Education (grade school, high school, college, graduate)

2. Estimation Difference in means or a model

3. Checking Determine matched sample size, tweak, repeat, ...

Method 3: Coarsened Exact Matching

1. Preprocess (Matching)
   - Temporarily coarsen X as much as you're willing
     - e.g., Education (grade school, high school, college, graduate)
     - Easy to understand, or can be automated as for a histogram

2. Estimation Difference in means or a model

3. Checking Determine matched sample size, tweak, repeat, ...

Method 3: Coarsened Exact Matching

1. Preprocess (Matching)
   - Temporarily coarsen X as much as you're willing
     - e.g., Education (grade school, high school, college, graduate)
     - Easy to understand, or can be automated as for a histogram
   - Apply exact matching to the coarsened X, $C(X)$

2. Estimation Difference in means or a model

3. Checking Determine matched sample size, tweak, repeat, ...

Method 3: Coarsened Exact Matching

1. Preprocess (Matching)
   - Temporarily coarsen X as much as you're willing
     - e.g., Education (grade school, high school, college, graduate)
     - Easy to understand, or can be automated as for a histogram
   - Apply exact matching to the coarsened X, $C(X)$
     - Sort observations into strata, each with unique values of $C(X)$

2. Estimation Difference in means or a model

3. Checking Determine matched sample size, tweak, repeat, ...

Method 3: Coarsened Exact Matching

1. Preprocess (Matching)
   - Temporarily coarsen X as much as you're willing
     - e.g., Education (grade school, high school, college, graduate)
     - Easy to understand, or can be automated as for a histogram
   - Apply exact matching to the coarsened X, $C(X)$
     - Sort observations into strata, each with unique values of $C(X)$
     - Prune any stratum with 0 treated or 0 control units

2. Estimation Difference in means or a model

3. Checking Determine matched sample size, tweak, repeat, ...

Method 3: Coarsened Exact Matching

1. Preprocess (Matching)
   - Temporarily coarsen X as much as you're willing
     - e.g., Education (grade school, high school, college, graduate)
     - Easy to understand, or can be automated as for a histogram
   - Apply exact matching to the coarsened X, $C(X)$
     - Sort observations into strata, each with unique values of $C(X)$
     - Prune any stratum with 0 treated or 0 control units
   - Pass on original (uncoarsened) units except those pruned
2. Estimation Difference in means or a model

3. Checking Determine matched sample size, tweak, repeat, ...

Method 3: Coarsened Exact Matching

1. Preprocess (Matching)
   - Temporarily coarsen X as much as you're willing
     - e.g., Education (grade school, high school, college, graduate)
     - Easy to understand, or can be automated as for a histogram
   - Apply exact matching to the coarsened X, $C(X)$
     - Sort observations into strata, each with unique values of $C(X)$
     - Prune any stratum with 0 treated or 0 control units
   - Pass on original (uncoarsened) units except those pruned
2. Estimation Difference in means or a model
   - Need to weight controls in each stratum to equal treateds

3. Checking Determine matched sample size, tweak, repeat, ...

Method 3: Coarsened Exact Matching

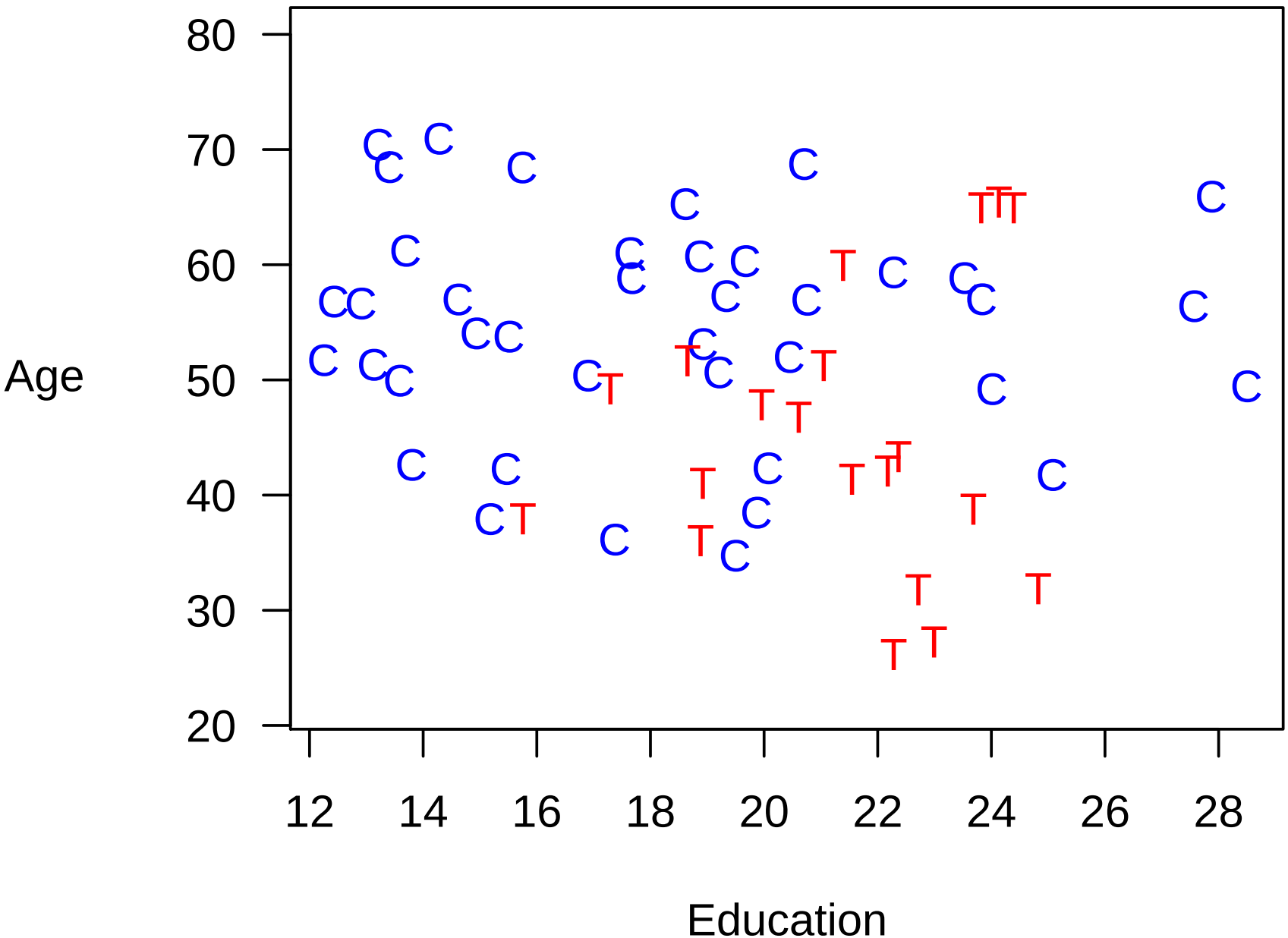
1. Preprocess (Matching)
   - Temporarily coarsen X as much as you're willing
     - e.g., Education (grade school, high school, college, graduate)
     - Easy to understand, or can be automated as for a histogram
   - Apply exact matching to the coarsened X, $C(X)$
     - Sort observations into strata, each with unique values of $C(X)$
     - Prune any stratum with 0 treated or 0 control units
   - Pass on original (uncoarsened) units except those pruned
2. Estimation Difference in means or a model
   - Need to weight controls in each stratum to equal treateds
   - Can apply other matching methods within CEM strata (inherit CEM's properties)
3. Checking Determine matched sample size, tweak, repeat, ...

Method 3: Coarsened Exact Matching

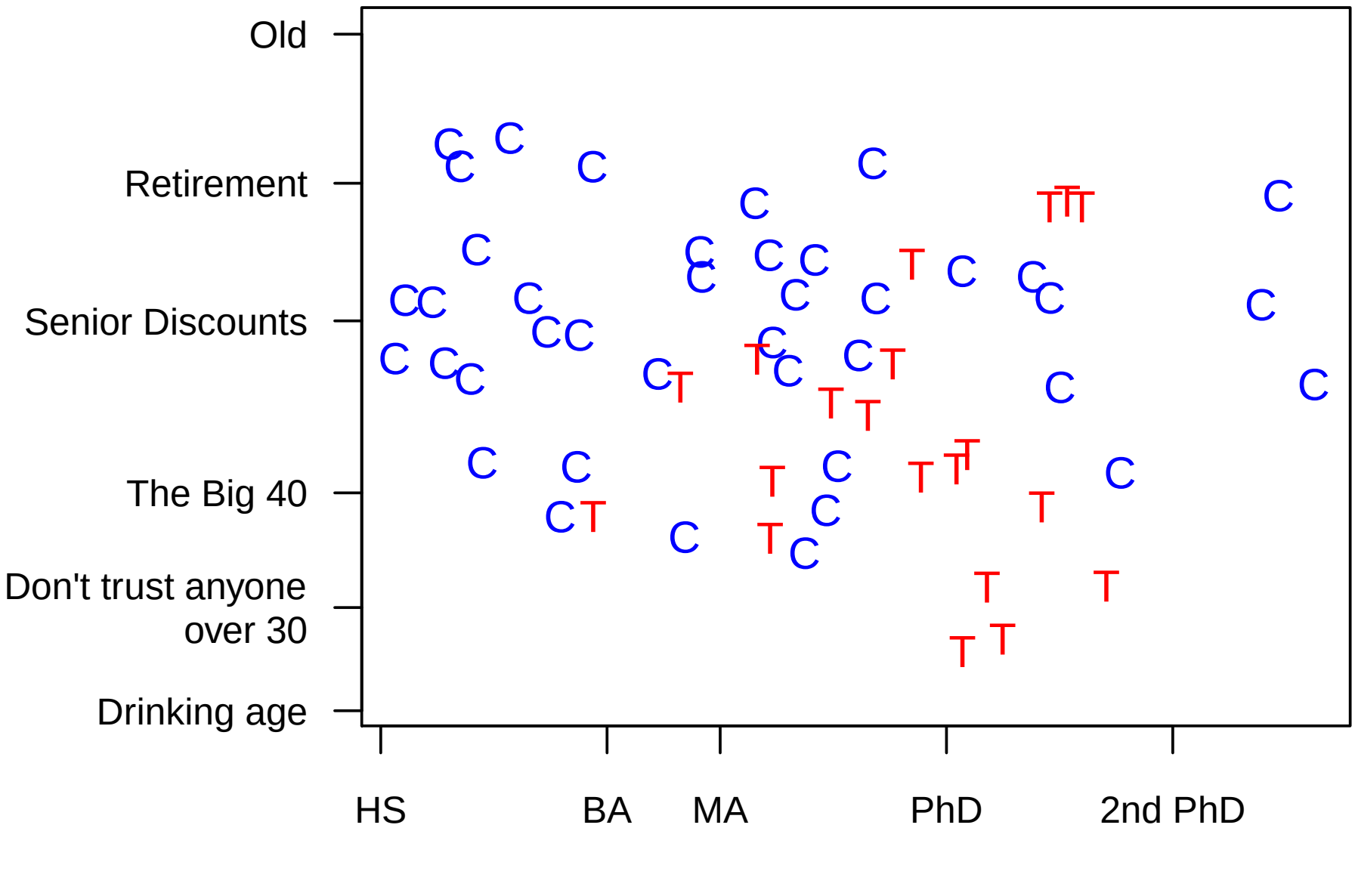
1. Preprocess (Matching)
   - Temporarily coarsen X as much as you're willing
     - e.g., Education (grade school, high school, college, graduate)
     - Easy to understand, or can be automated as for a histogram
   - Apply exact matching to the coarsened X, $C(X)$
     - Sort observations into strata, each with unique values of $C(X)$
     - Prune any stratum with 0 treated or 0 control units
   - Pass on original (uncoarsened) units except those pruned
2. Estimation Difference in means or a model
   - Need to weight controls in each stratum to equal treateds
   - Can apply other matching methods within CEM strata (inherit CEM's properties)
3. Checking Determine matched sample size, tweak, repeat, ...
   - Easier, but still iterative

Coarsened Exact Matching

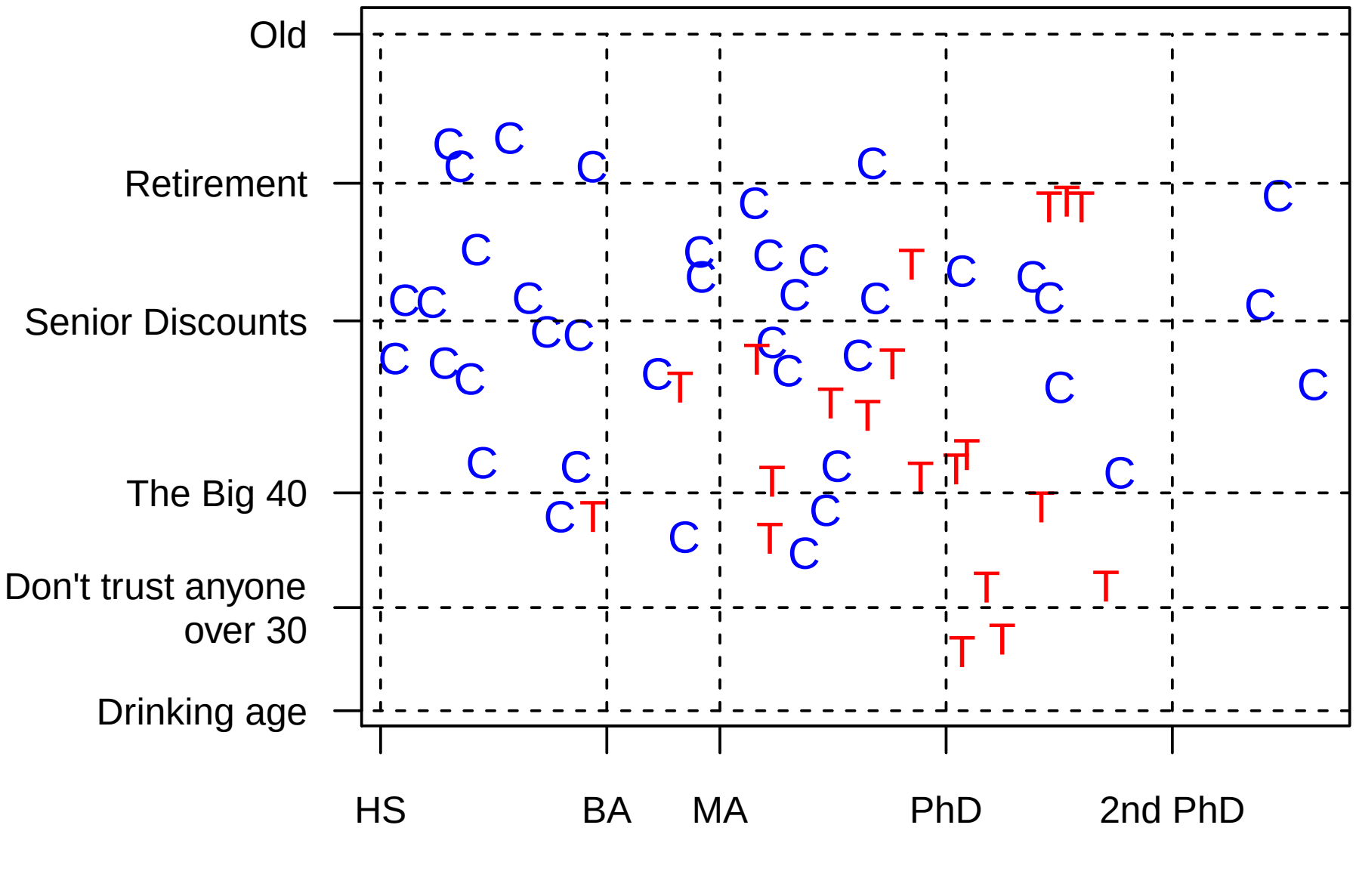
Coarsened Exact Matching



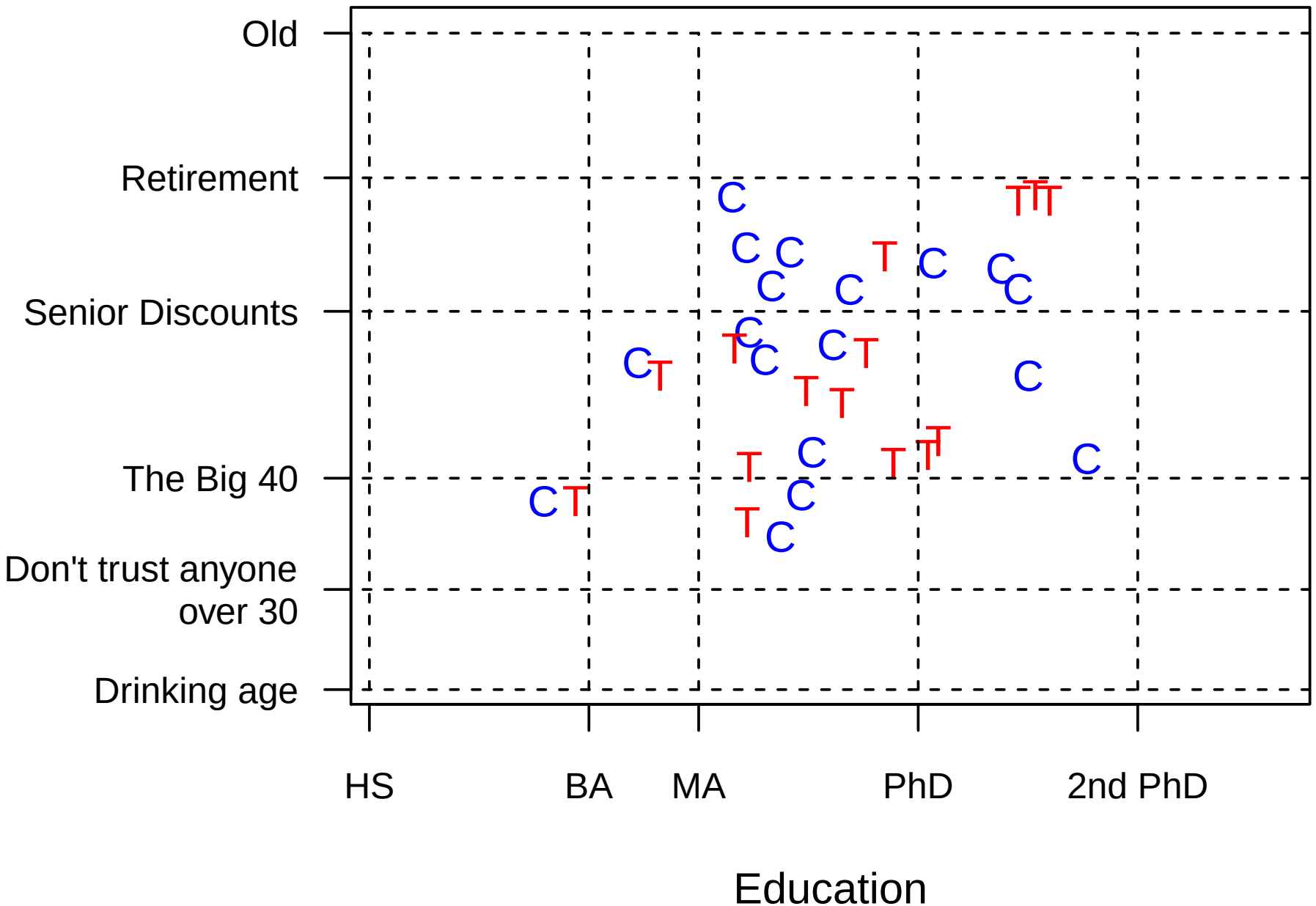
Coarsened Exact Matching



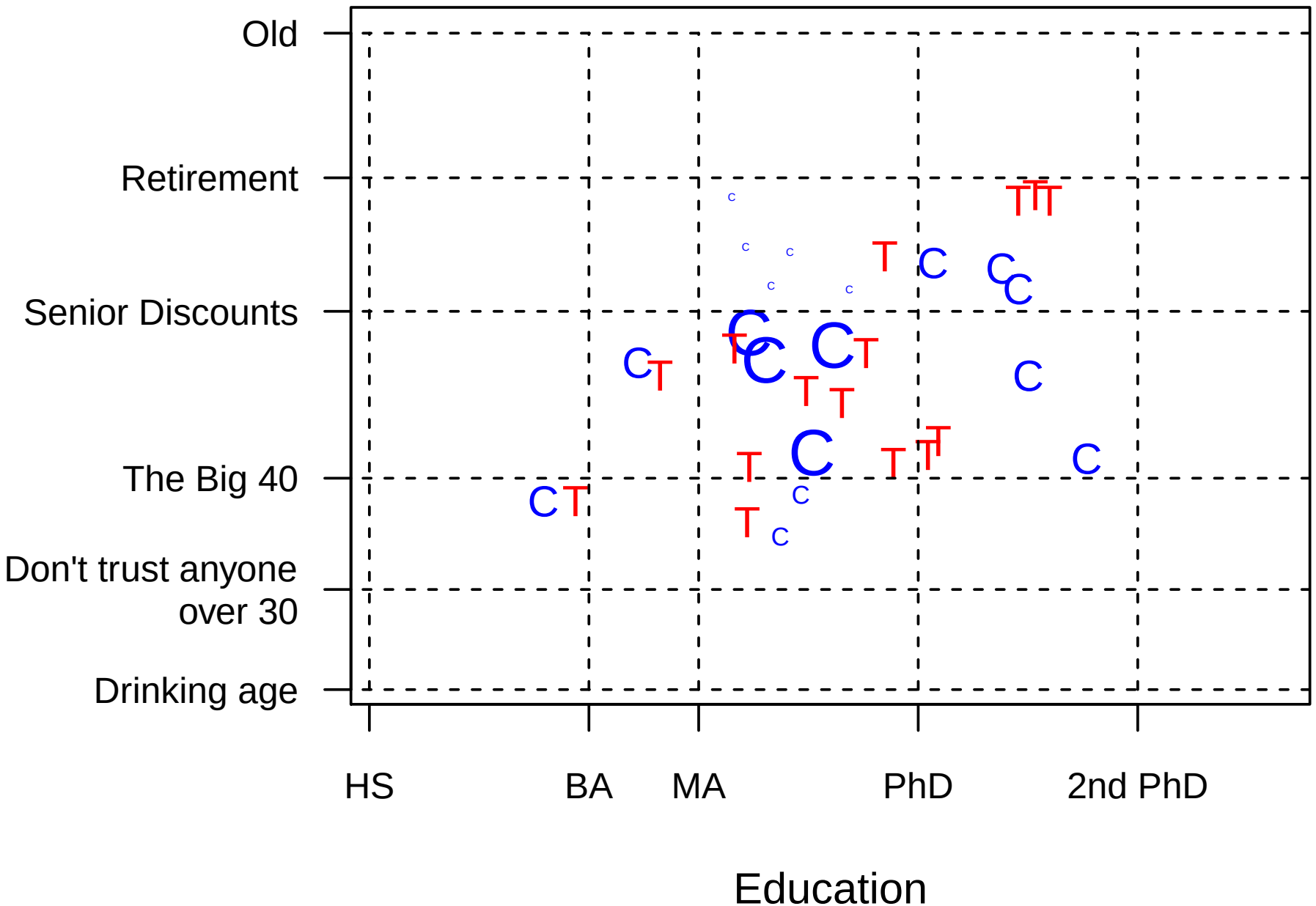
Coarsened Exact Matching



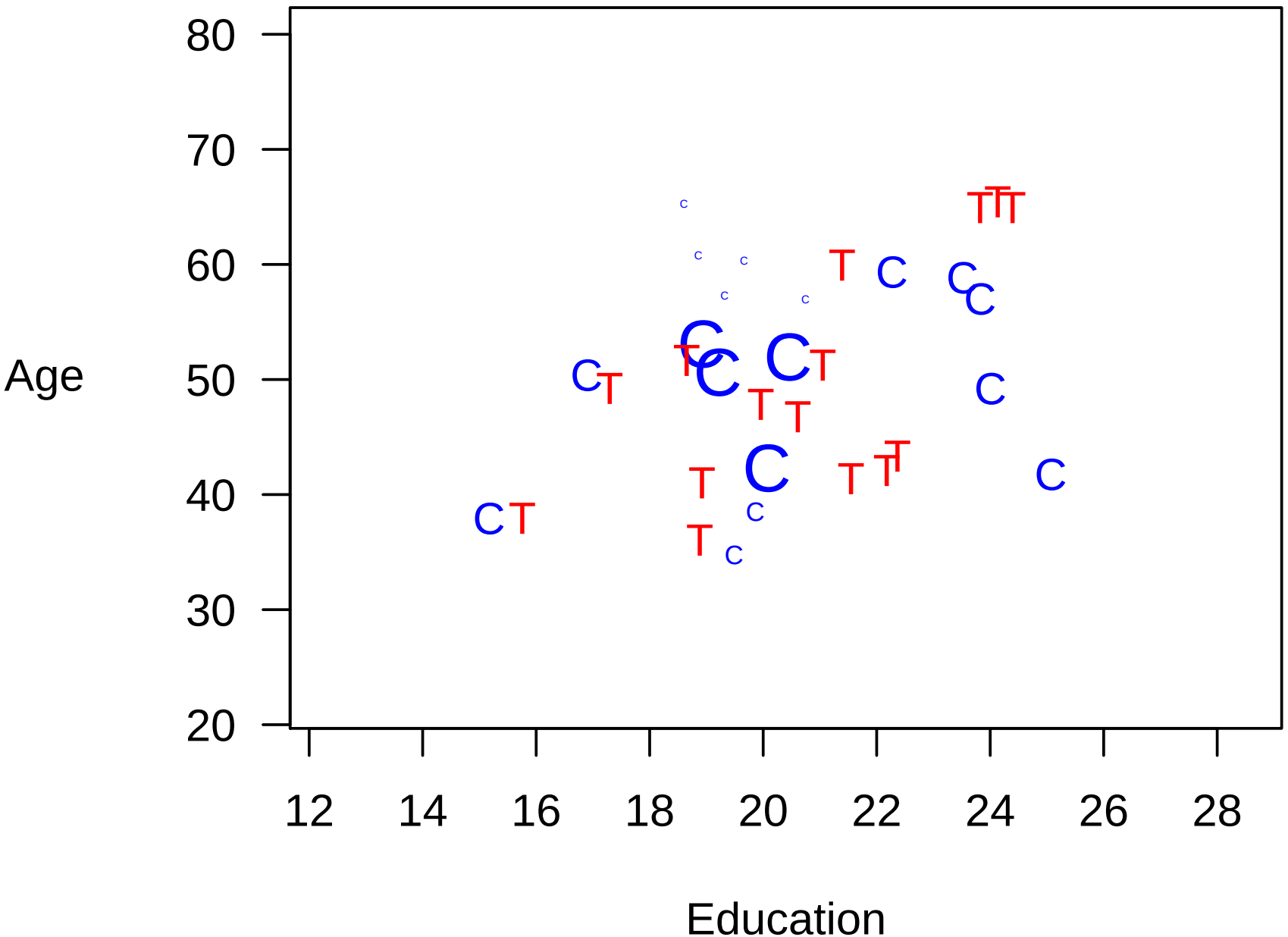
Coarsened Exact Matching



Coarsened Exact Matching



Coarsened Exact Matching



The Matching Frontier

The Matching Frontier

- Bias-Variance trade off $\rightsquigarrow$ Imbalance-n Trade Off
  Frontier = matched dataset with lowest imbalance for each n

The Matching Frontier

- Bias-Variance trade off $\rightsquigarrow$ Imbalance-n Trade Off
  Frontier = matched dataset with lowest imbalance for each n
- To use, make 3 choices:

The Matching Frontier

- Bias-Variance trade off $\rightsquigarrow$ Imbalance-n Trade Off
  Frontier = matched dataset with lowest imbalance for each n
- To use, make 3 choices:
  1. Imbalance metric, e.g.:

The Matching Frontier

- Bias-Variance trade off $\rightsquigarrow$ Imbalance-n Trade Off
  Frontier = matched dataset with lowest imbalance for each n
- To use, make 3 choices:
  1. Imbalance metric, e.g.:
     - Average Mahalanobis Distance (average distance from each unit to the closest in the other treatment regime)

The Matching Frontier

- Bias-Variance trade off $\rightsquigarrow$ Imbalance-n Trade Off
  Frontier = matched dataset with lowest imbalance for each n
- To use, make 3 choices:
  1. Imbalance metric, e.g.:
     - Average Mahalanobis Distance (average distance from each unit to the closest in the other treatment regime)
     - Difference of multivariate histograms (L_1):

The Matching Frontier

- Bias-Variance trade off $\rightsquigarrow$ Imbalance-n Trade Off
  Frontier = matched dataset with lowest imbalance for each n
- To use, make 3 choices:
  1. Imbalance metric, e.g.:
     - Average Mahalanobis Distance (average distance from each unit to the closest in the other treatment regime)
     - Difference of multivariate histograms (L_1):
  2. Quantity of interest: SATT (prune Cs only) or FSATT

The Matching Frontier

- Bias-Variance trade off $\rightsquigarrow$ Imbalance-n Trade Off
  Frontier = matched dataset with lowest imbalance for each n
- To use, make 3 choices:
  1. Imbalance metric, e.g.:
     - Average Mahalanobis Distance (average distance from each unit to the closest in the other treatment regime)
     - Difference of multivariate histograms (L_1):
  2. Quantity of interest: SATT (prune Cs only) or FSATT
  3. Fixed- or variable-ratio matching

The Matching Frontier

- Bias-Variance trade off $\rightsquigarrow$ Imbalance-n Trade Off
  Frontier = matched dataset with lowest imbalance for each n
- To use, make 3 choices:
  1. Imbalance metric, e.g.:
     - Average Mahalanobis Distance (average distance from each unit to the closest in the other treatment regime)
     - Difference of multivariate histograms (L_1):
  2. Quantity of interest: SATT (prune Cs only) or FSATT
  3. Fixed- or variable-ratio matching
- Result:

The Matching Frontier

- Bias-Variance trade off $\rightsquigarrow$ Imbalance-n Trade Off
  Frontier = matched dataset with lowest imbalance for each n
- To use, make 3 choices:
  1. Imbalance metric, e.g.:
     - Average Mahalanobis Distance (average distance from each unit to the closest in the other treatment regime)
     - Difference of multivariate histograms (L_1):
  2. Quantity of interest: SATT (prune Cs only) or FSATT
  3. Fixed- or variable-ratio matching
- Result:
  - Simple to use

The Matching Frontier

- Bias-Variance trade off $\rightsquigarrow$ Imbalance-n Trade Off
  Frontier = matched dataset with lowest imbalance for each n
- To use, make 3 choices:
  1. Imbalance metric, e.g.:
     - Average Mahalanobis Distance (average distance from each unit to the closest in the other treatment regime)
     - Difference of multivariate histograms (L_1):
  2. Quantity of interest: SATT (prune Cs only) or FSATT
  3. Fixed- or variable-ratio matching
- Result:
  - Simple to use
  - All solutions are optimal

The Matching Frontier

- Bias-Variance trade off $\rightsquigarrow$ Imbalance-n Trade Off
  Frontier = matched dataset with lowest imbalance for each n
- To use, make 3 choices:
  1. Imbalance metric, e.g.:
     - Average Mahalanobis Distance (average distance from each unit to the closest in the other treatment regime)
     - Difference of multivariate histograms (L_1):
  2. Quantity of interest: SATT (prune Cs only) or FSATT
  3. Fixed- or variable-ratio matching
- Result:
  - Simple to use
  - All solutions are optimal
  - No iteration or diagnostics required

The Matching Frontier

- Bias-Variance trade off $\rightsquigarrow$ Imbalance-n Trade Off
  Frontier = matched dataset with lowest imbalance for each n
- To use, make 3 choices:
  1. Imbalance metric, e.g.:
     - Average Mahalanobis Distance (average distance from each unit to the closest in the other treatment regime)
     - Difference of multivariate histograms (L_1):
  2. Quantity of interest: SATT (prune Cs only) or FSATT
  3. Fixed- or variable-ratio matching
- Result:
  - Simple to use
  - All solutions are optimal
  - No iteration or diagnostics required
  - No cherry picking possible

How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:

How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with $(N \times k)$ control matrix X_0

How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with $(N \times k)$ control matrix X_0
  - Calculate imbalance for all $\binom{N}{n}$ subsets of rows of X_0

How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with $(N \times k)$ control matrix X_0
  - Calculate imbalance for all $\binom{N}{n}$ subsets of rows of X_0
  - Choose the (or a) subset with the lowest imbalance

How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with $(N \times k)$ control matrix X_0
  - Calculate imbalance for all $\binom{N}{n}$ subsets of rows of X_0
  - Choose the (or a) subset with the lowest imbalance
- Evaluations needed to compute the entire frontier:

How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with $(N \times k)$ control matrix X_0
  - Calculate imbalance for all $\binom{N}{n}$ subsets of rows of X_0
  - Choose the (or a) subset with the lowest imbalance
- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$ evaluations for each sample size $n = N, N-1, \ldots, 1$

How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with $(N \times k)$ control matrix X_0
  - Calculate imbalance for all $\binom{N}{n}$ subsets of rows of X_0
  - Choose the (or a) subset with the lowest imbalance
- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$ evaluations for each sample size $n = N, N-1, \ldots, 1$
  - The combination is the (gargantuan) "power set"

How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with $(N \times k)$ control matrix X_0
  - Calculate imbalance for all $\binom{N}{n}$ subsets of rows of X_0
  - Choose the (or a) subset with the lowest imbalance
- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$ evaluations for each sample size $n = N, N-1, \ldots, 1$
  - The combination is the (gargantuan) "power set"
  - e.g., $N > 300$ requires more imbalance evaluations than elementary particles in the universe!

How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with $(N \times k)$ control matrix X_0
  - Calculate imbalance for all $\binom{N}{n}$ subsets of rows of X_0
  - Choose the (or a) subset with the lowest imbalance
- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$ evaluations for each sample size $n = N, N-1, \ldots, 1$
  - The combination is the (gargantuan) "power set"
  - e.g., $N > 300$ requires more imbalance evaluations than elementary particles in the universe!
  - $\rightsquigarrow$ It's **hard** to calculate!

How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with $(N \times k)$ control matrix X_0
  - Calculate imbalance for all $\binom{N}{n}$ subsets of rows of X_0
  - Choose the (or a) subset with the lowest imbalance
- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$ evaluations for each sample size $n = N, N-1, \ldots, 1$
  - The combination is the (gargantuan) "power set"
  - e.g., $N > 300$ requires more imbalance evaluations than elementary particles in the universe!
  - $\rightsquigarrow$ It's **hard** to calculate!
- We develop new algorithms for several frontiers which:

How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with $(N \times k)$ control matrix X_0
  - Calculate imbalance for all $\binom{N}{n}$ subsets of rows of X_0
  - Choose the (or a) subset with the lowest imbalance
- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$ evaluations for each sample size $n = N, N-1, \ldots, 1$
  - The combination is the (gargantuan) "power set"
  - e.g., $N > 300$ requires more imbalance evaluations than elementary particles in the universe!
  - $\rightsquigarrow$ It's **hard** to calculate!
- We develop new algorithms for several frontiers which:
  - run very fast

How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with $(N \times k)$ control matrix X_0
  - Calculate imbalance for all $\binom{N}{n}$ subsets of rows of X_0
  - Choose the (or a) subset with the lowest imbalance
- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$ evaluations for each sample size $n = N, N-1, \ldots, 1$
  - The combination is the (gargantuan) "power set"
  - e.g., $N > 300$ requires more imbalance evaluations than elementary particles in the universe!
  - $\rightsquigarrow$ It's **hard** to calculate!
- We develop new algorithms for several frontiers which:
  - run very fast
  - do not require evaluating every subset

How hard is the frontier to calculate?

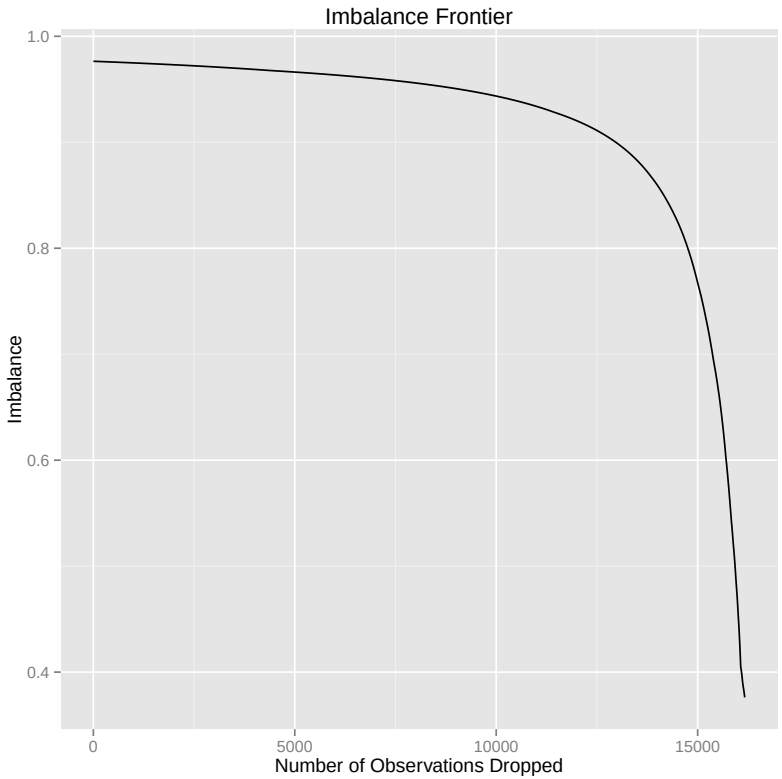
- Consider 1 point on the SATT frontier:
  - Start with $(N \times k)$ control matrix X_0
  - Calculate imbalance for all $\binom{N}{n}$ subsets of rows of X_0
  - Choose the (or a) subset with the lowest imbalance
- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$ evaluations for each sample size $n = N, N-1, \ldots, 1$
  - The combination is the (gargantuan) "power set"
  - e.g., $N > 300$ requires more imbalance evaluations than elementary particles in the universe!
  - $\rightsquigarrow$ It's **hard** to calculate!
- We develop new algorithms for several frontiers which:
  - run very fast
  - do not require evaluating every subset
  - work with very large data sets

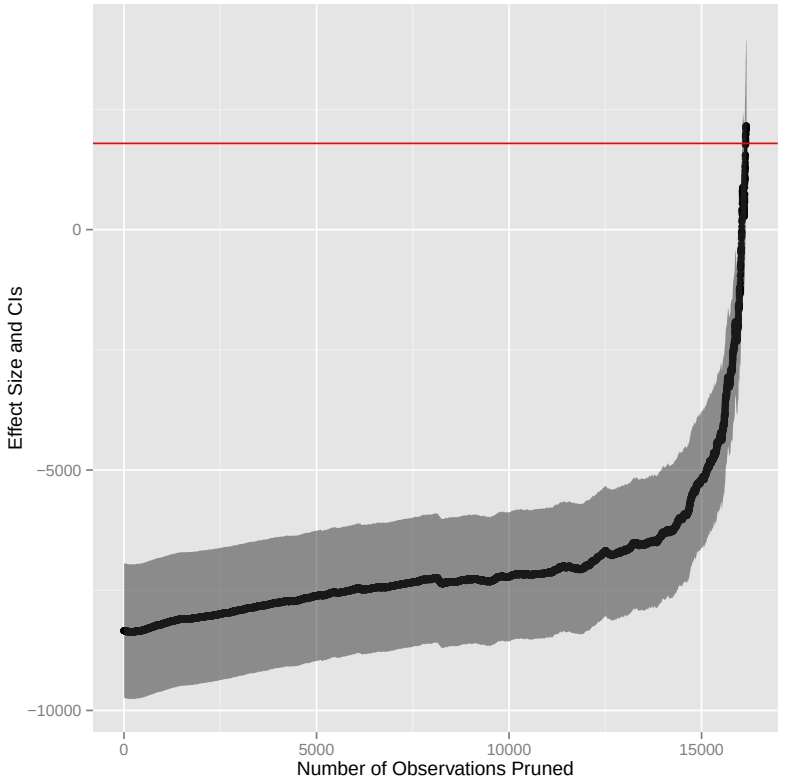
How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with $(N \times k)$ control matrix X_0
  - Calculate imbalance for all $\binom{N}{n}$ subsets of rows of X_0
  - Choose the (or a) subset with the lowest imbalance
- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$ evaluations for each sample size $n = N, N-1, \ldots, 1$
  - The combination is the (gargantuan) "power set"
  - e.g., $N > 300$ requires more imbalance evaluations than elementary particles in the universe!
  - $\rightsquigarrow$ It's **hard** to calculate!
- We develop new algorithms for several frontiers which:
  - run very fast
  - do not require evaluating every subset
  - work with very large data sets
  - $\rightsquigarrow$ It's **easy** to calculate!

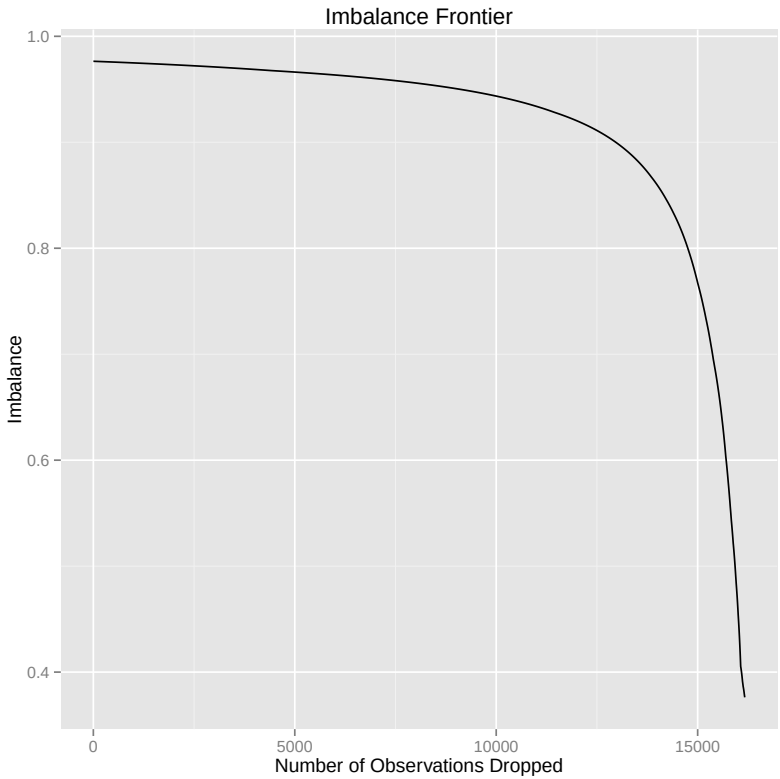
Job Training Data: Frontier and Causal Estimates

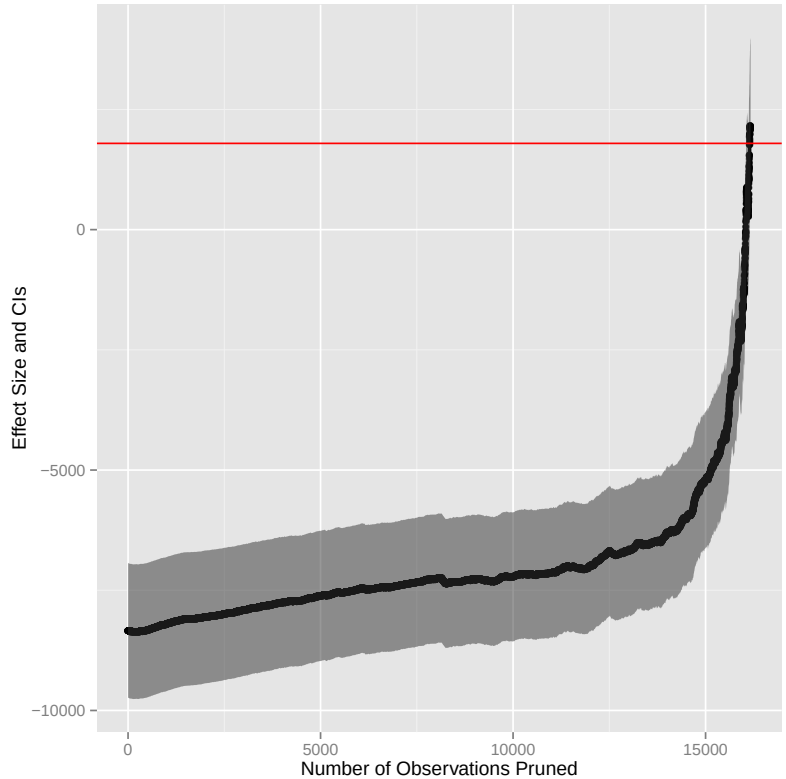
Job Training Data: Frontier and Causal Estimates





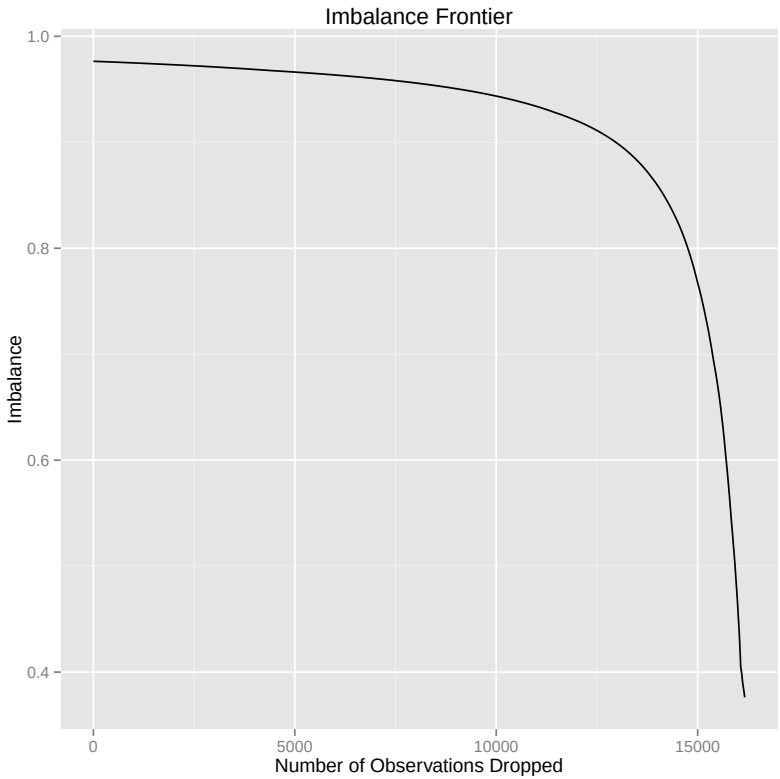
Job Training Data: Frontier and Causal Estimates

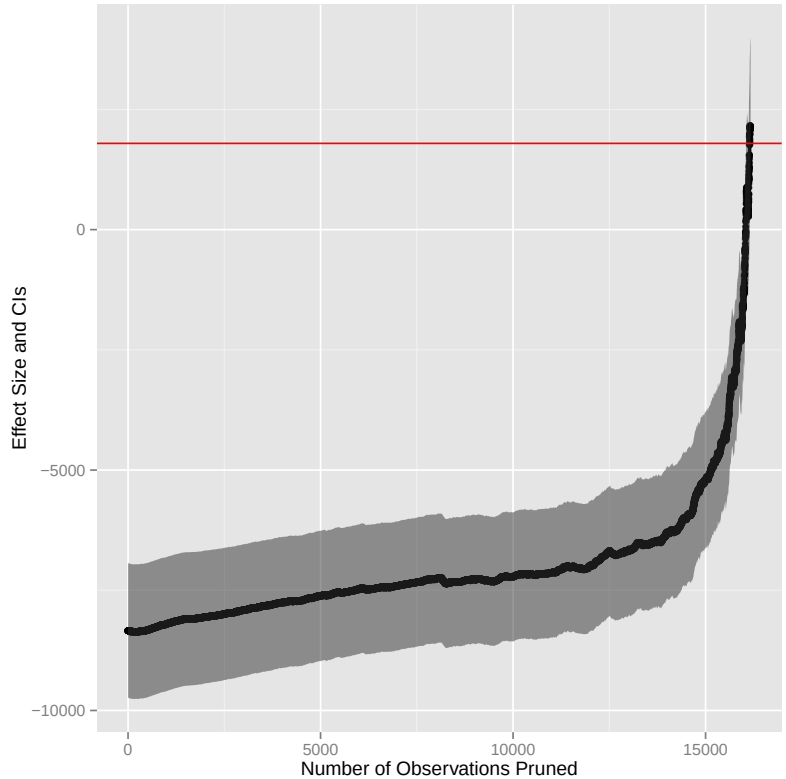




- 185 Ts; pruning most 16,252 Cs won't increase variance much

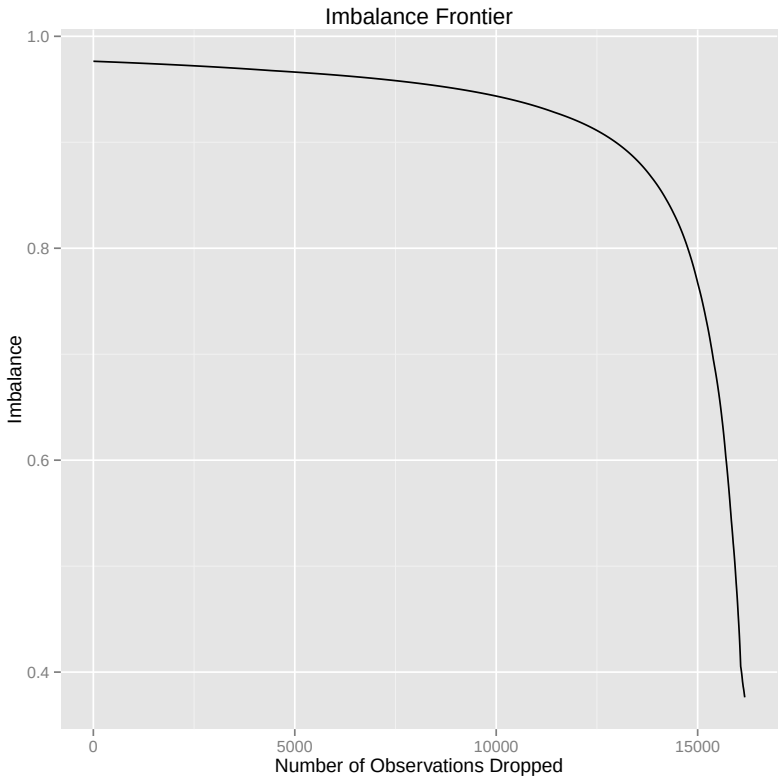
Job Training Data: Frontier and Causal Estimates

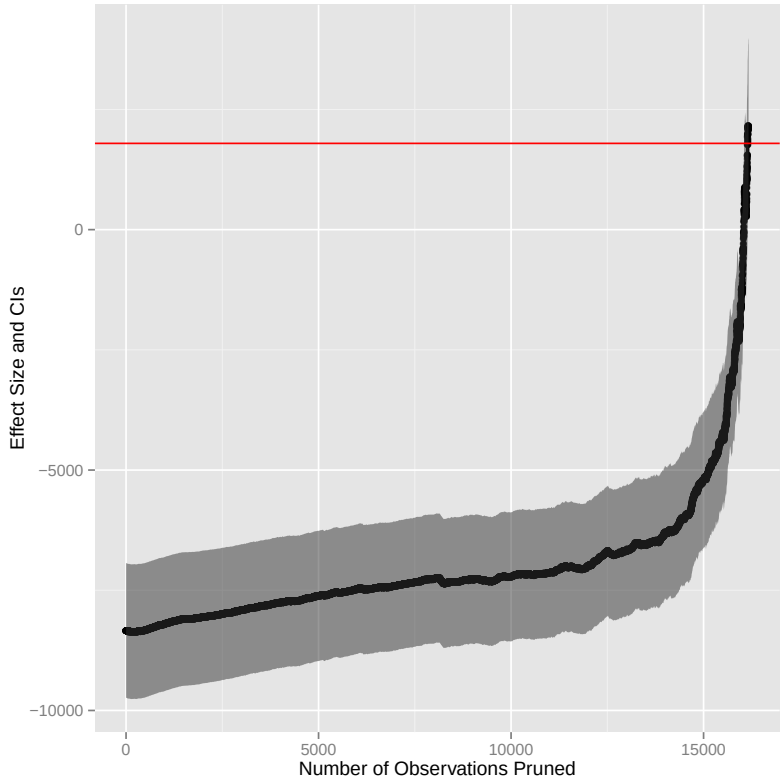




- 185 Ts; pruning most 16,252 Cs won't increase variance much
- Huge bias-variance trade-off after most are pruned

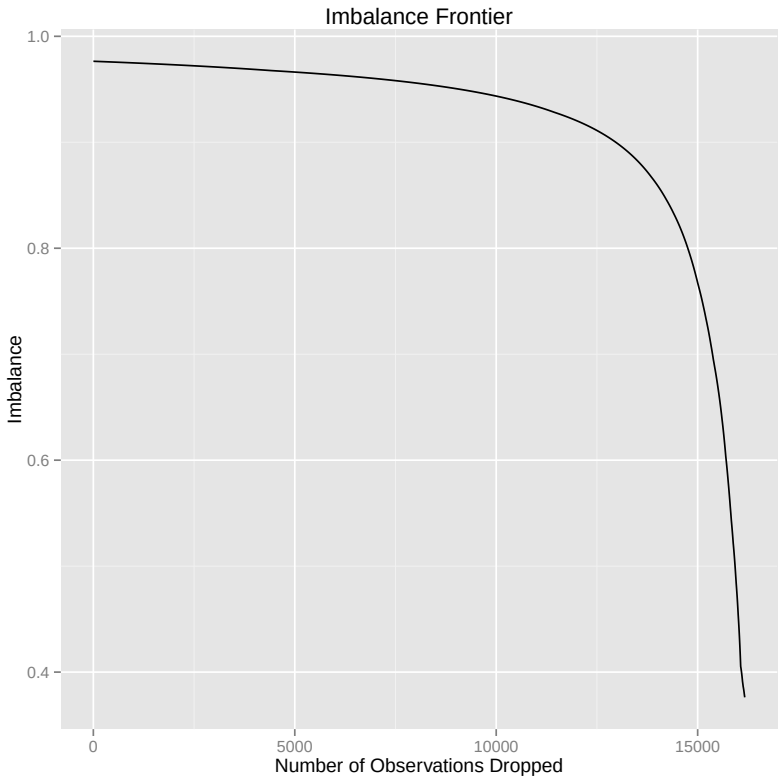
Job Training Data: Frontier and Causal Estimates

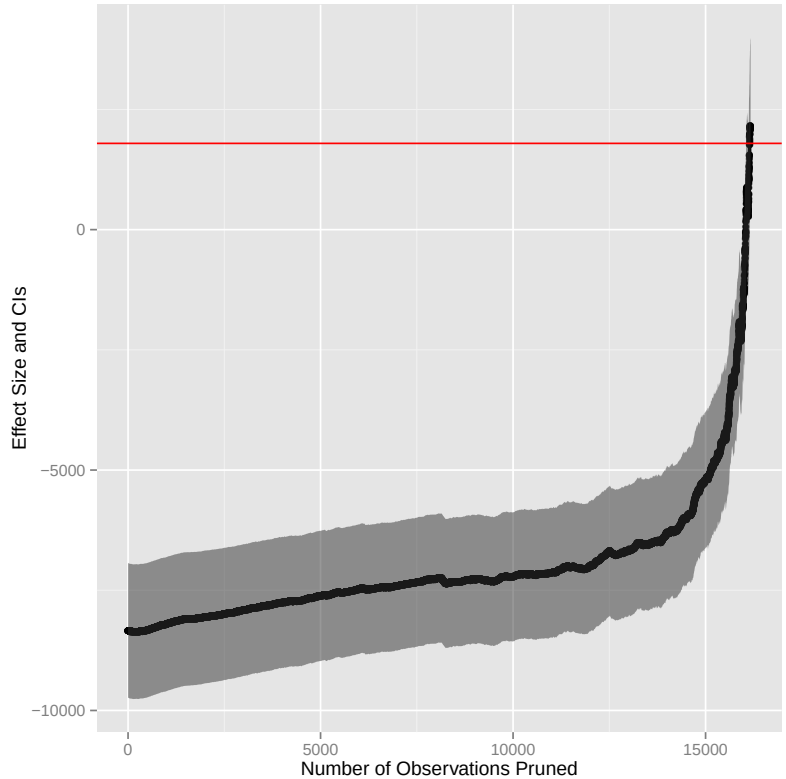




- 185 Ts; pruning most 16,252 Cs won't increase variance much
- Huge bias-variance trade-off after most are pruned
- Estimates converge to experiment after removing bias

Job Training Data: Frontier and Causal Estimates

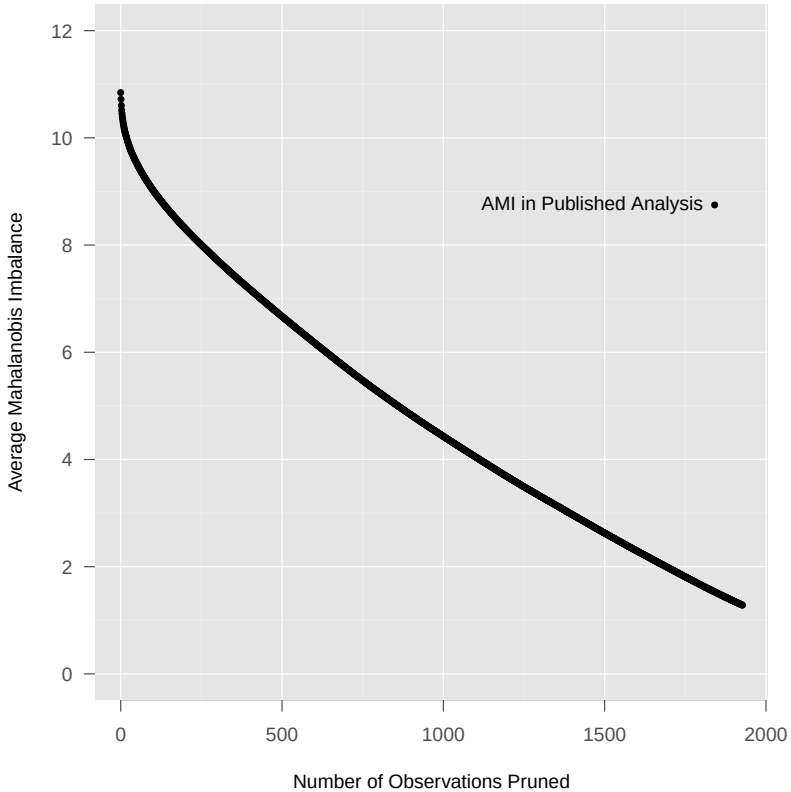


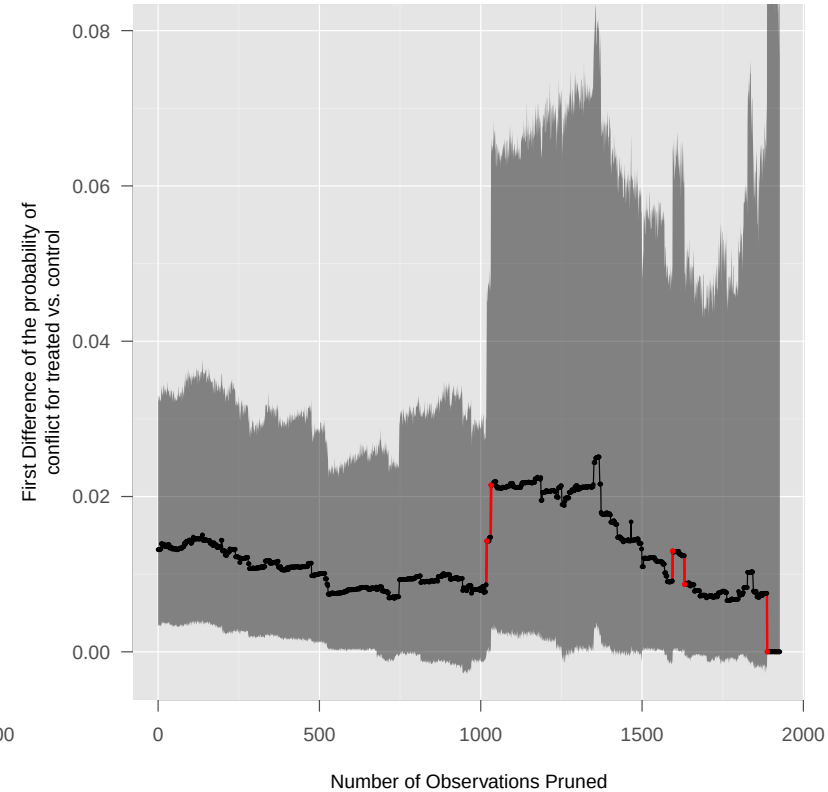


- 185 Ts; pruning most 16,252 Cs won't increase variance much
- Huge bias-variance trade-off after most are pruned
- Estimates converge to experiment after removing bias
- No mysteries: basis of inference clearly revealed

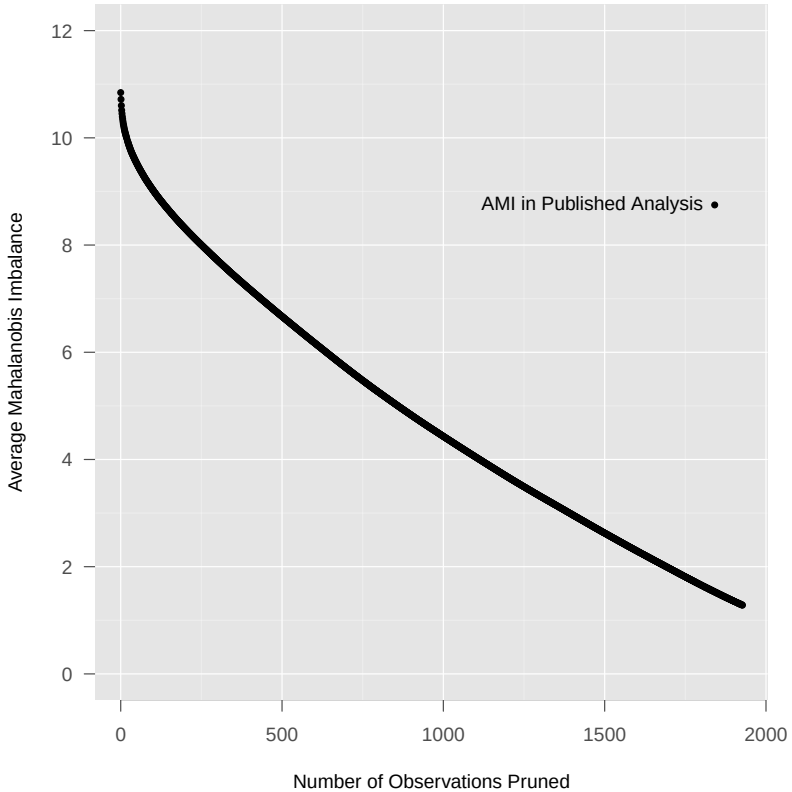
Aid Shocks Data: Frontier and Causal Estimates

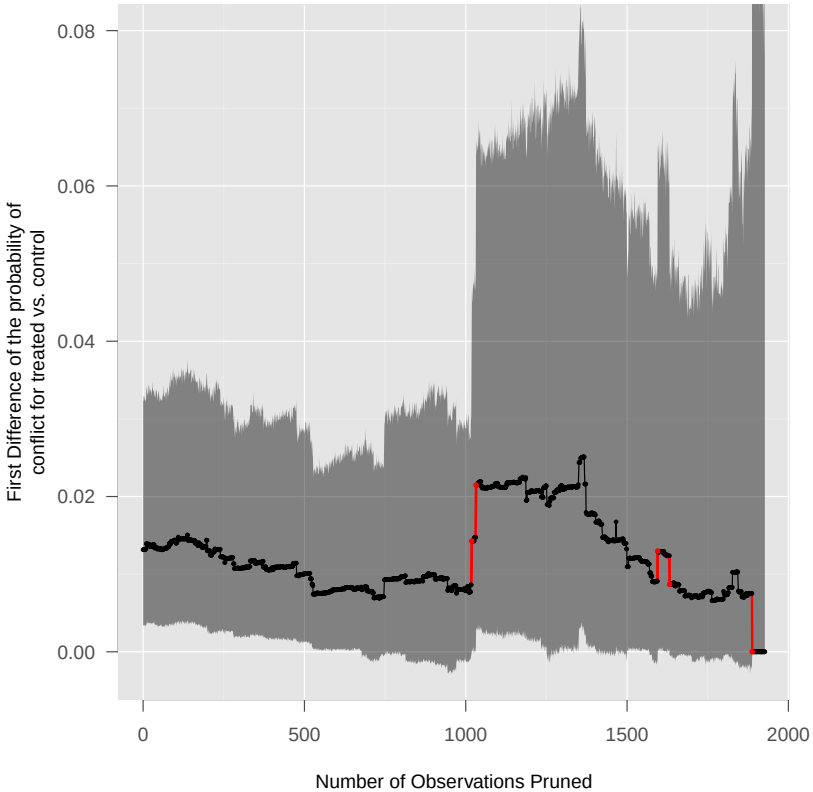
Aid Shocks Data: Frontier and Causal Estimates





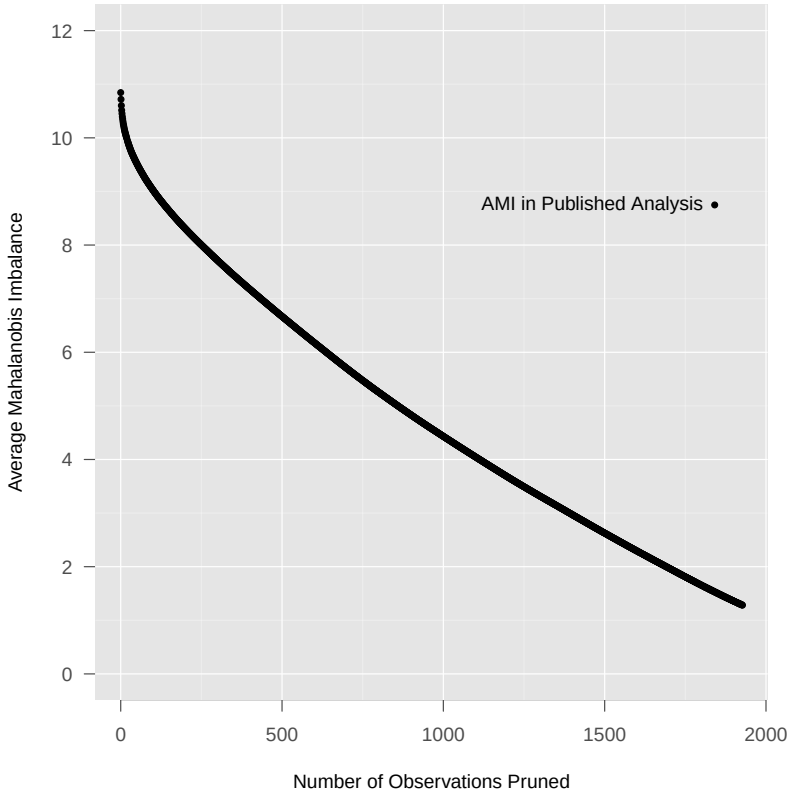
Aid Shocks Data: Frontier and Causal Estimates

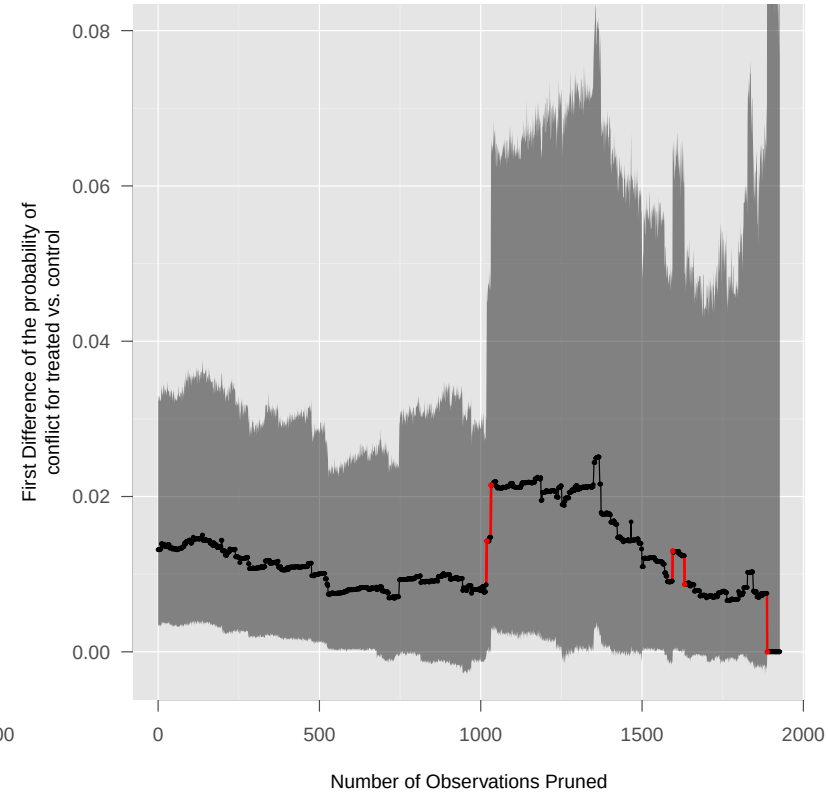




- Frontier is nearly linear (left)

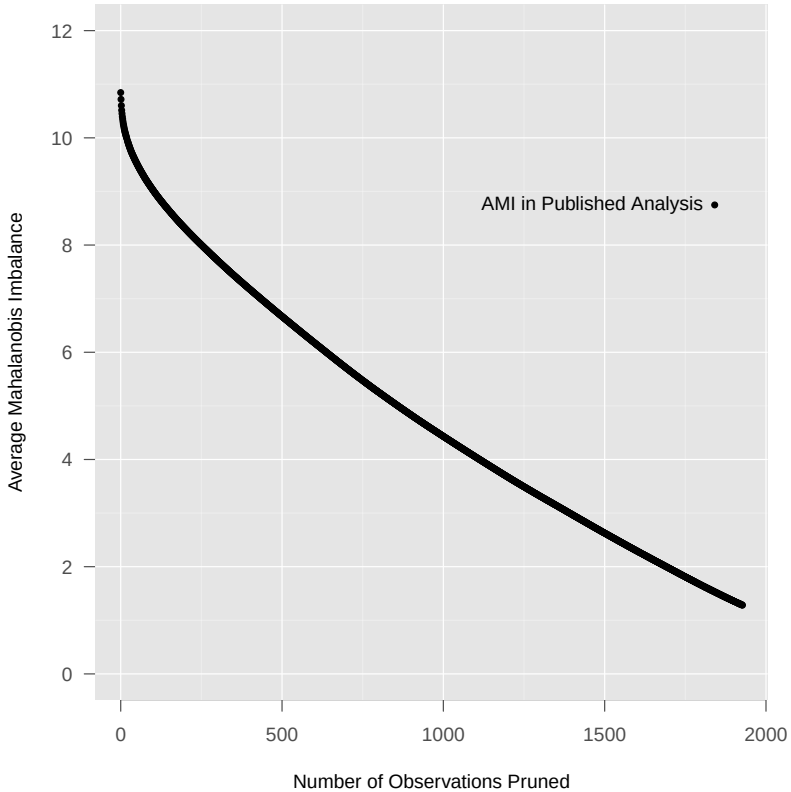
Aid Shocks Data: Frontier and Causal Estimates

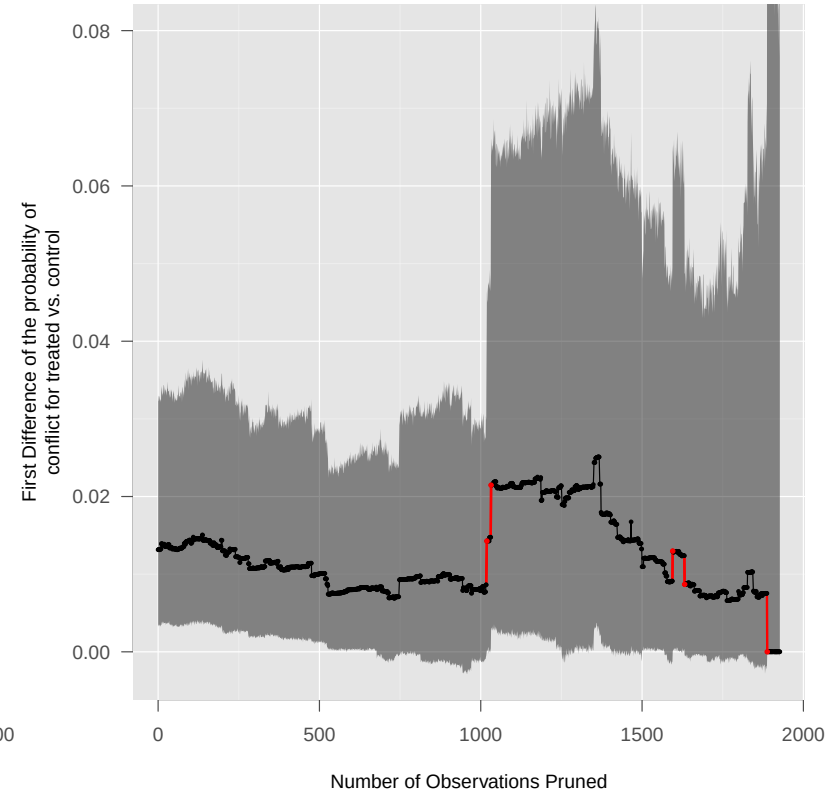




- Frontier is nearly linear (left)
- Causal effects have big jumps (right)

Aid Shocks Data: Frontier and Causal Estimates

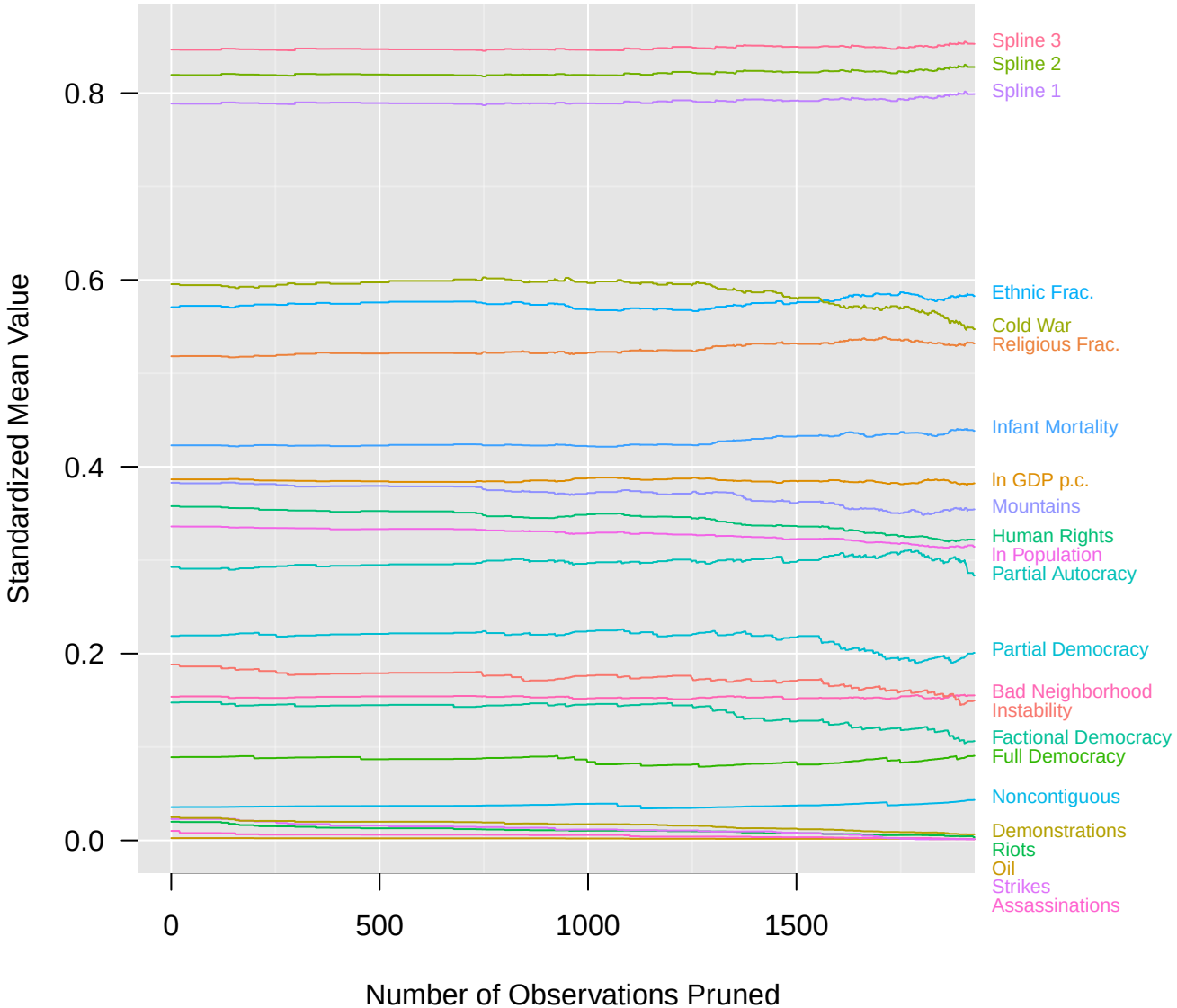




- Frontier is nearly linear (left)
- Causal effects have big jumps (right)
- More difficult inferential task

Aids Shocks: Change in Quantity of Interest

Aids Shocks: Change in Quantity of Interest



Aids Shocks: Large Unit-Level Effects

Aids Shocks: Large Unit-Level Effects

Case	T	Y	Effect change	N remaining
Gambia, 1991	1	0	0.008→0.015	1608
Niger, 1994	0	1	0.015→0.023	1595
Lesotho, 1998	1	1	0.021→0.018	1254
Cote D'Ivoire, 2002	1	1	0.011→0.008	995
Guinea, 2000	1	1	0.005→0	739

Aids Shocks: Large Unit-Level Effects

Case	T	Y	Effect change	N remaining
Gambia, 1991	1	0	0.008→0.015	1608
Niger, 1994	0	1	0.015→0.023	1595
Lesotho, 1998	1	1	0.021→0.018	1254
Cote D'Ivoire, 2002	1	1	0.011→0.008	995
Guinea, 2000	1	1	0.005→0	739

- High leverage points

Aids Shocks: Large Unit-Level Effects

Case	T	Y	Effect change	N remaining
Gambia, 1991	1	0	0.008→0.015	1608
Niger, 1994	0	1	0.015→0.023	1595
Lesotho, 1998	1	1	0.021→0.018	1254
Cote D'Ivoire, 2002	1	1	0.011→0.008	995
Guinea, 2000	1	1	0.005→0	739

- High leverage points
- Cases with few substitutes

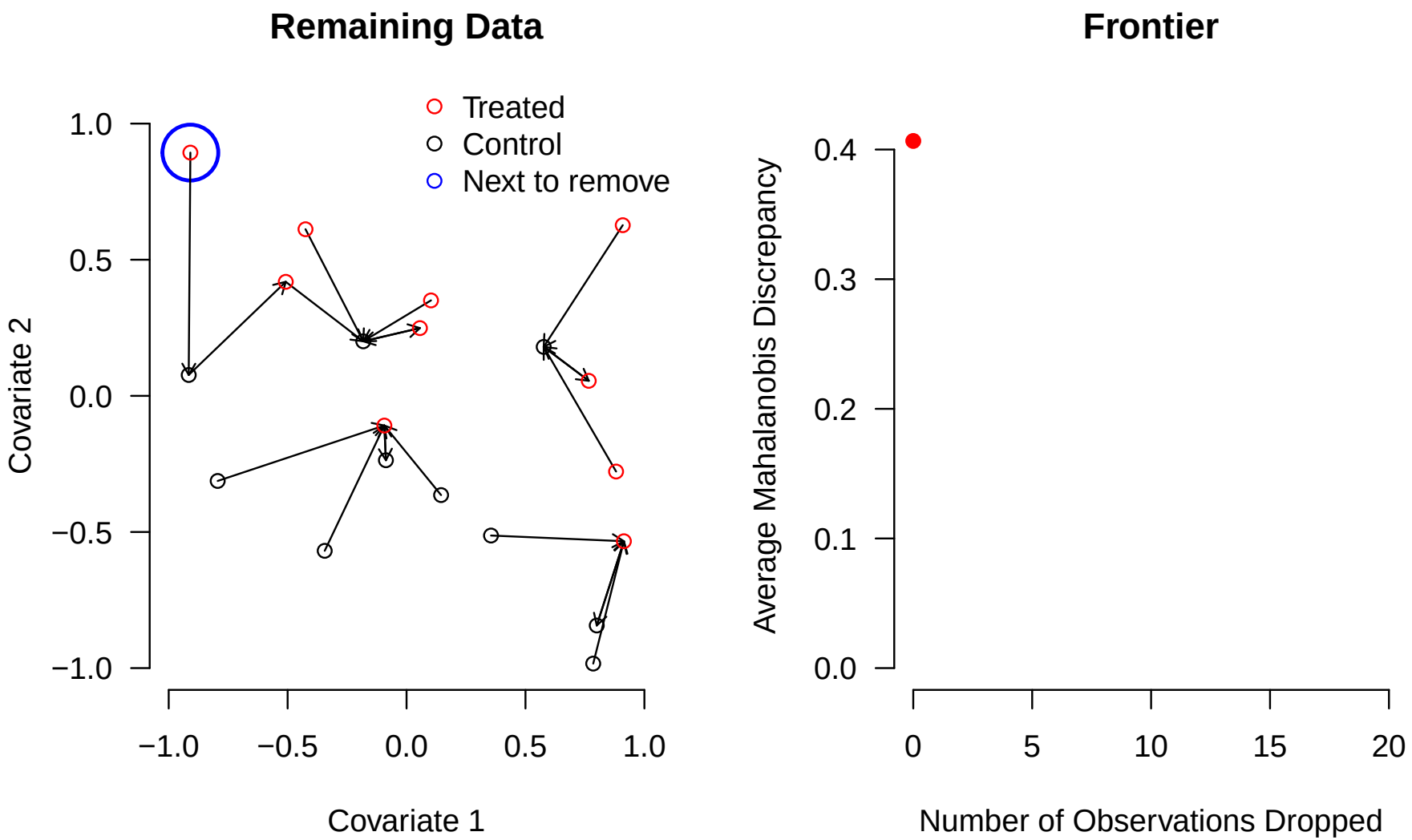
Aids Shocks: Large Unit-Level Effects

Case	T	Y	Effect change	N remaining
Gambia, 1991	1	0	0.008→0.015	1608
Niger, 1994	0	1	0.015→0.023	1595
Lesotho, 1998	1	1	0.021→0.018	1254
Cote D'Ivoire, 2002	1	1	0.011→0.008	995
Guinea, 2000	1	1	0.005→0	739

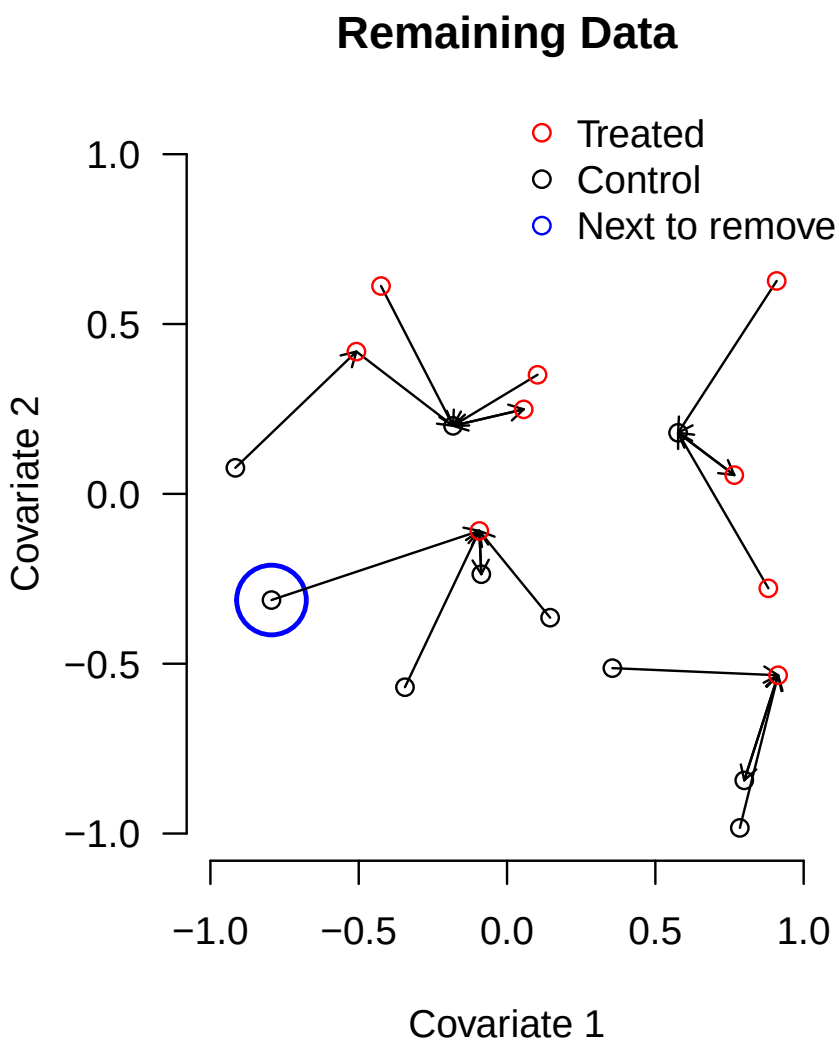
- High leverage points
- Cases with few substitutes
- Not model dependence (which matching helps with), but data dependence

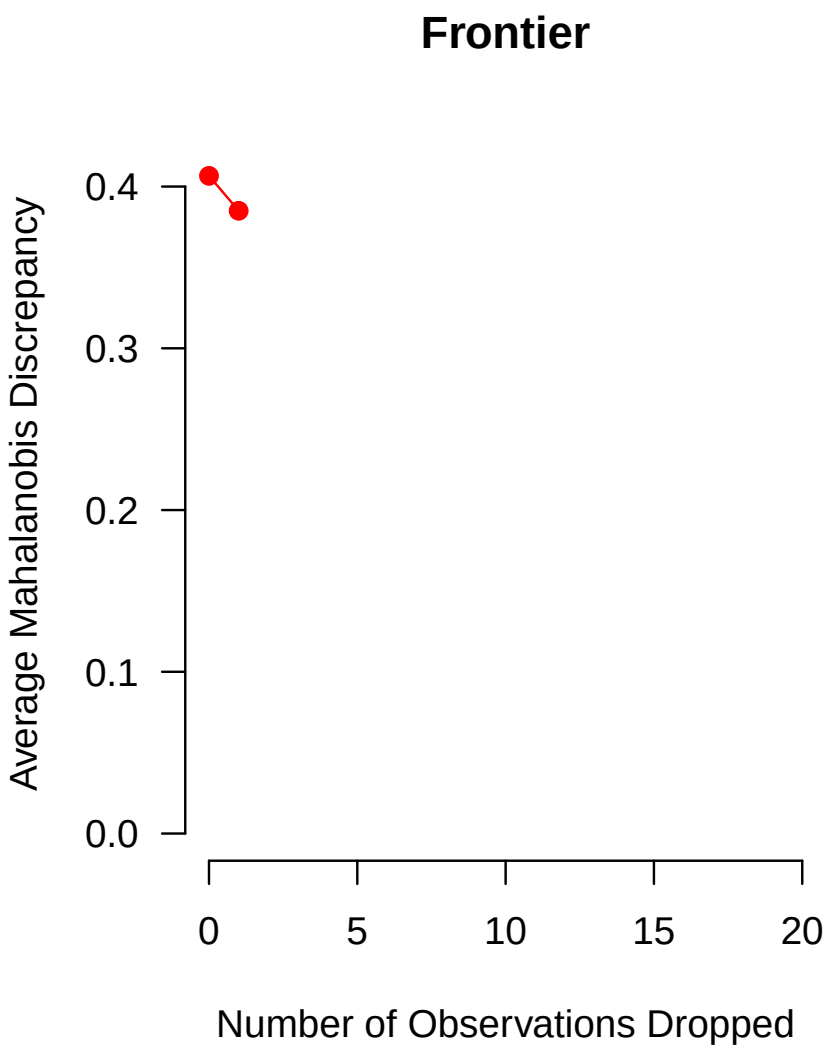
Constructing the FSATT Mahalanobis Frontier

Constructing the FSATT Mahalanobis Frontier

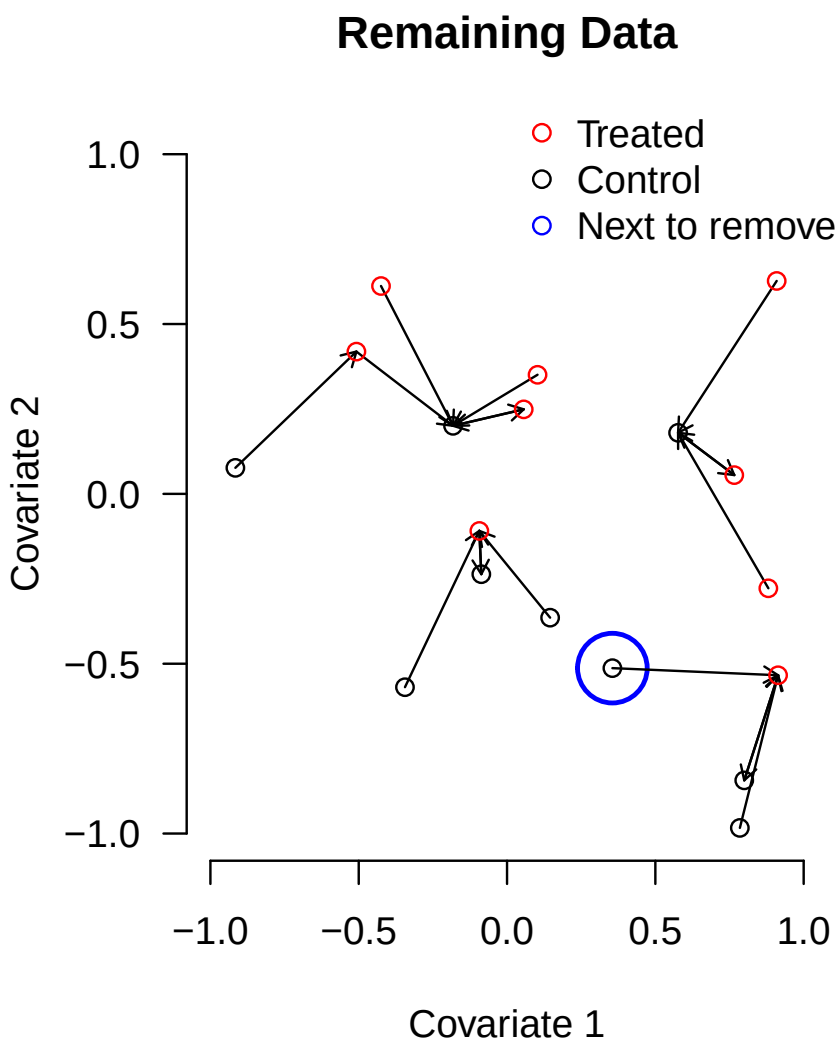


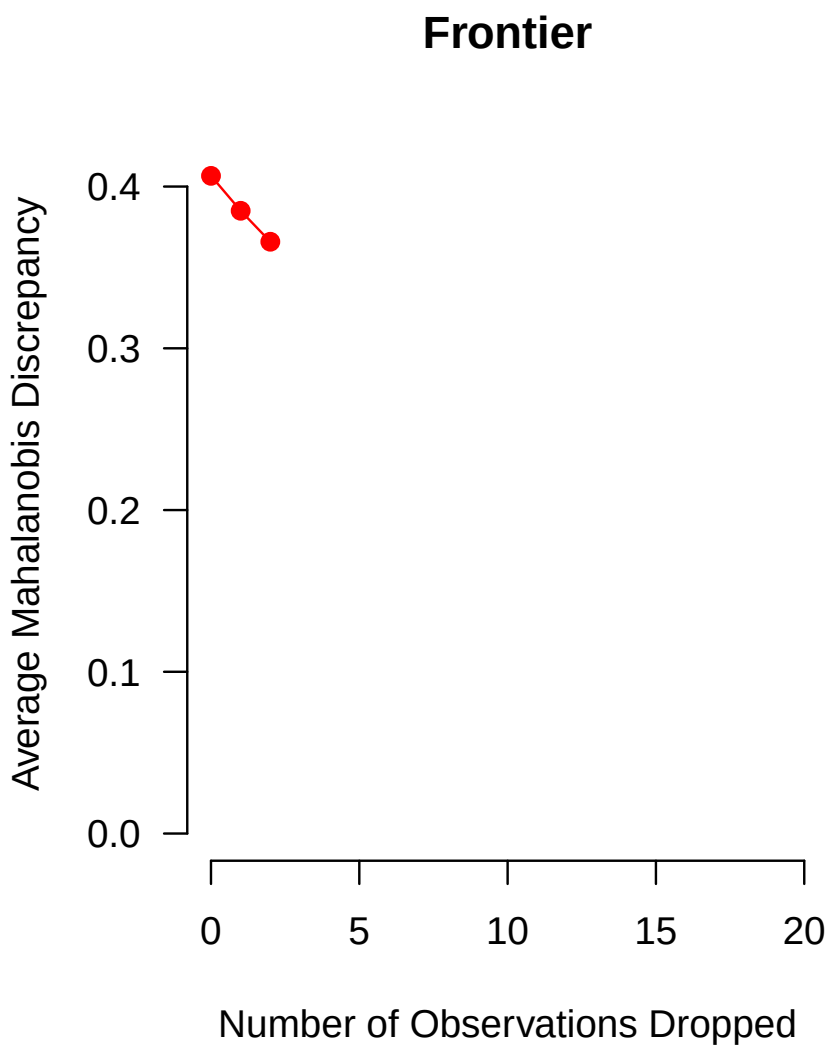
Constructing the FSATT Mahalanobis Frontier



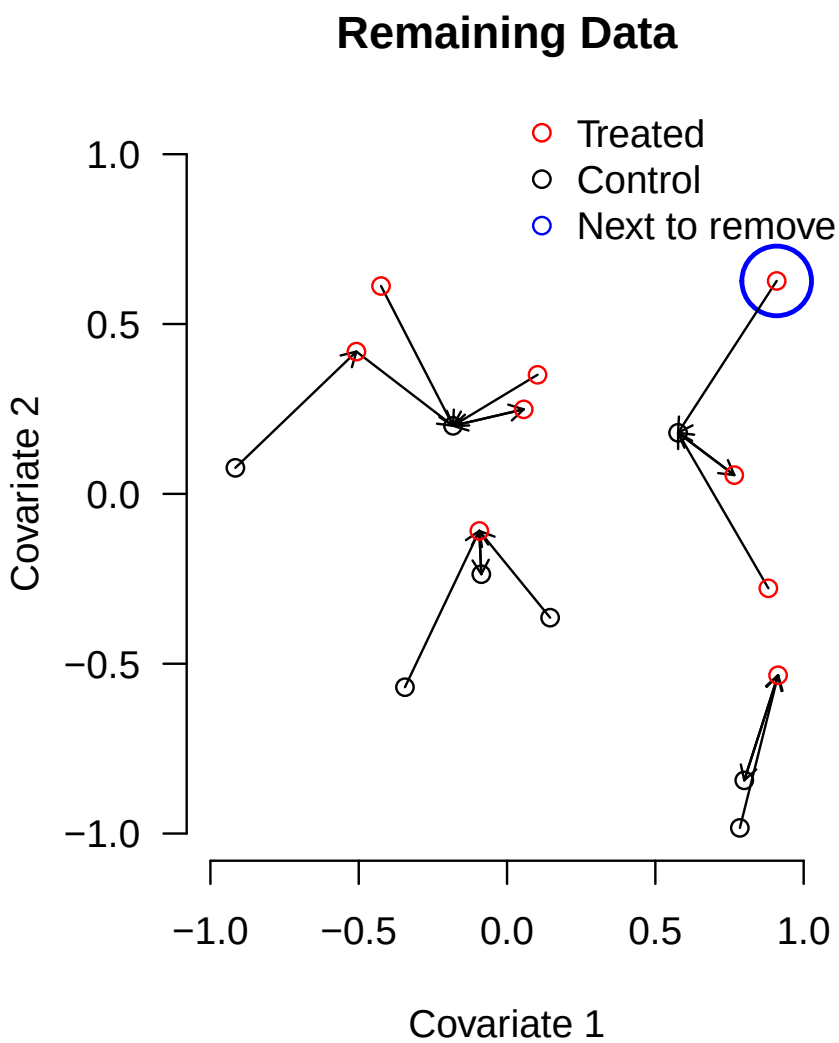


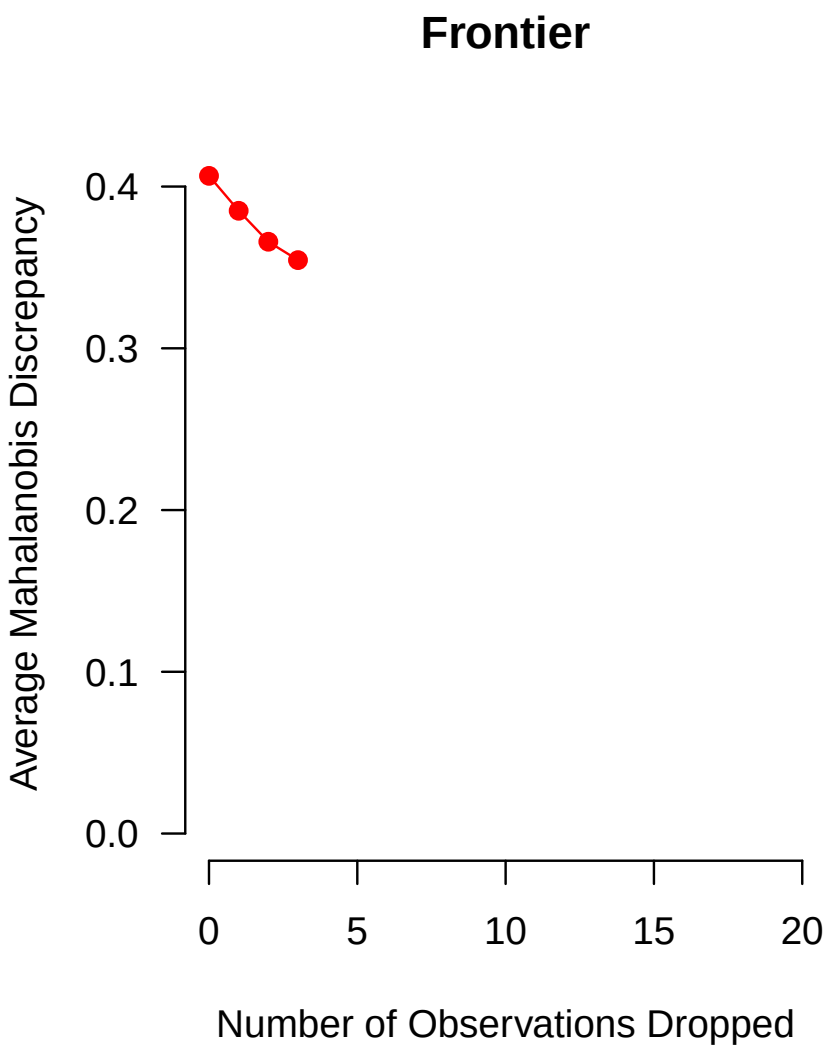
Constructing the FSATT Mahalanobis Frontier



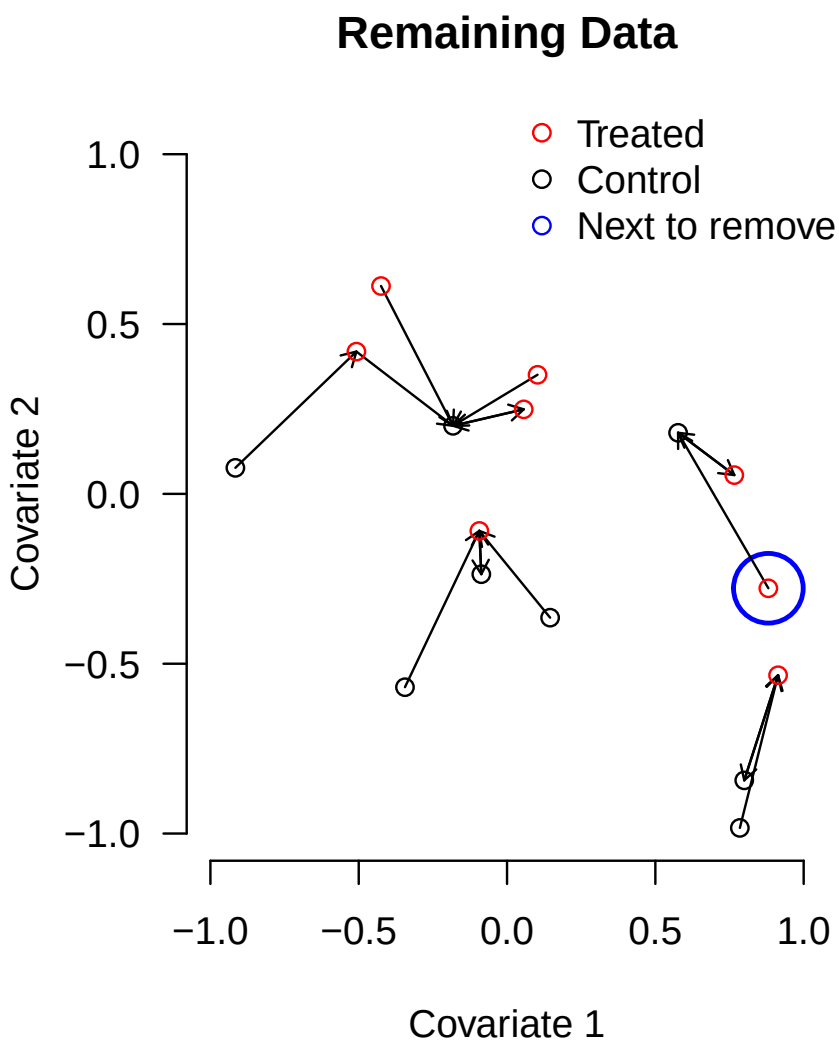


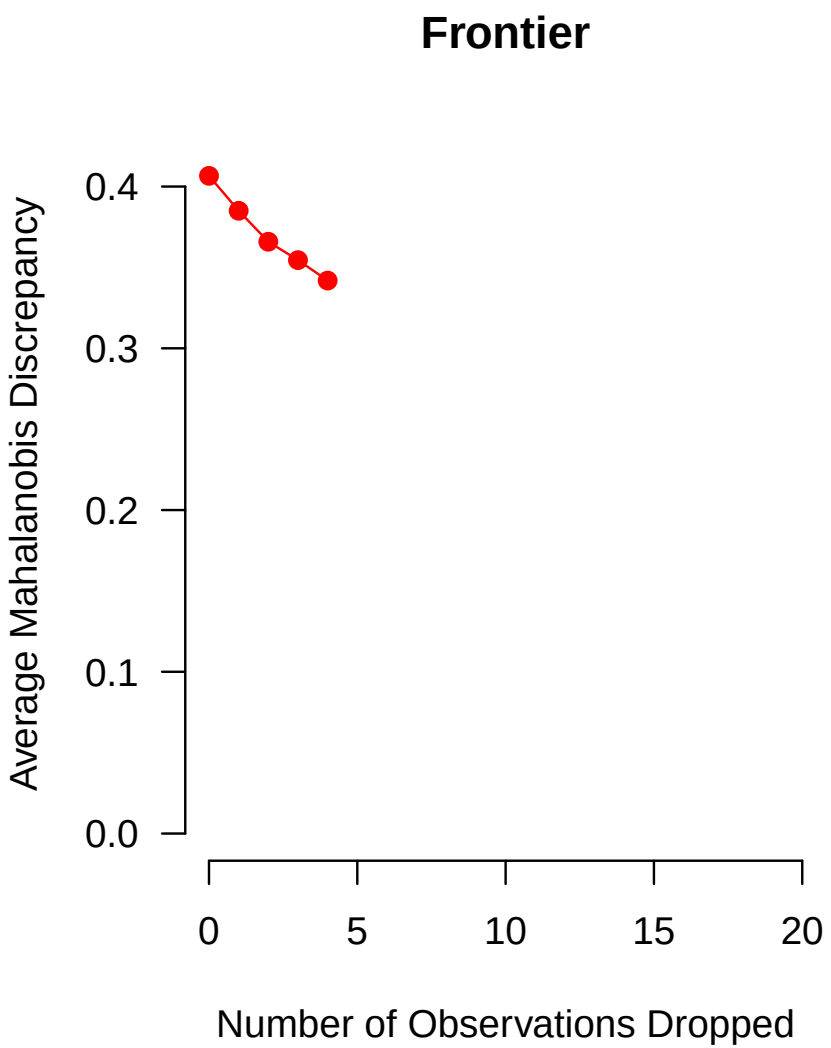
Constructing the FSATT Mahalanobis Frontier



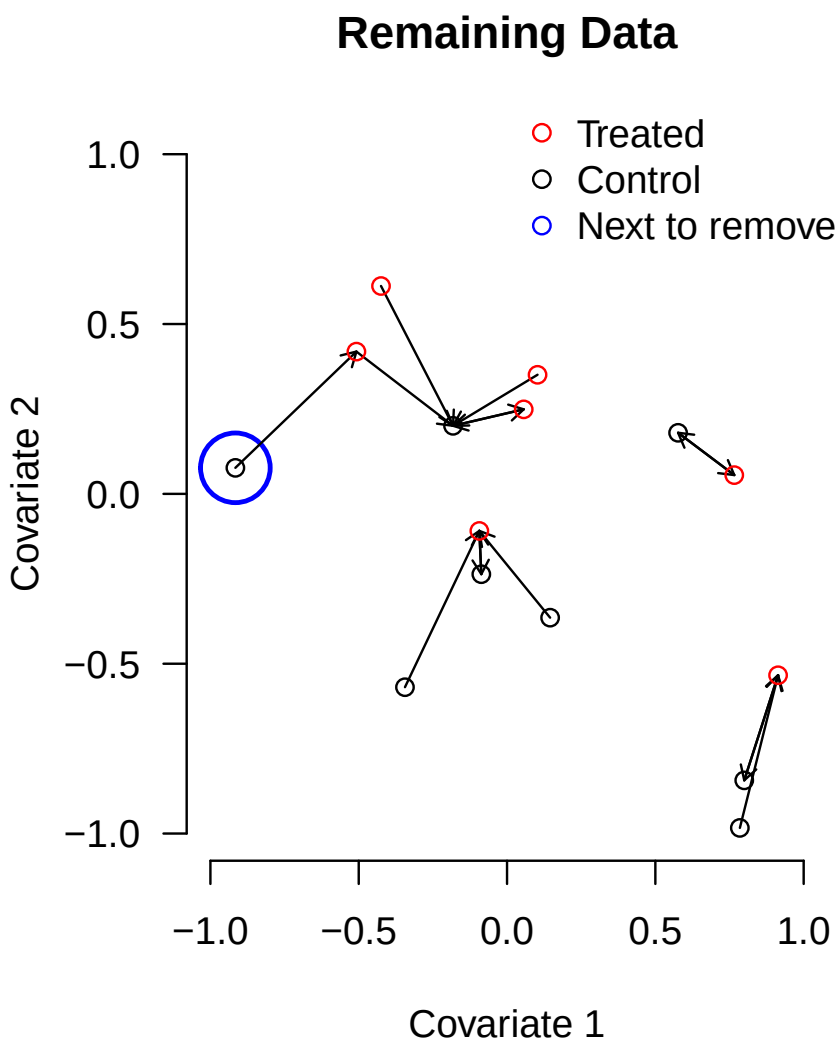


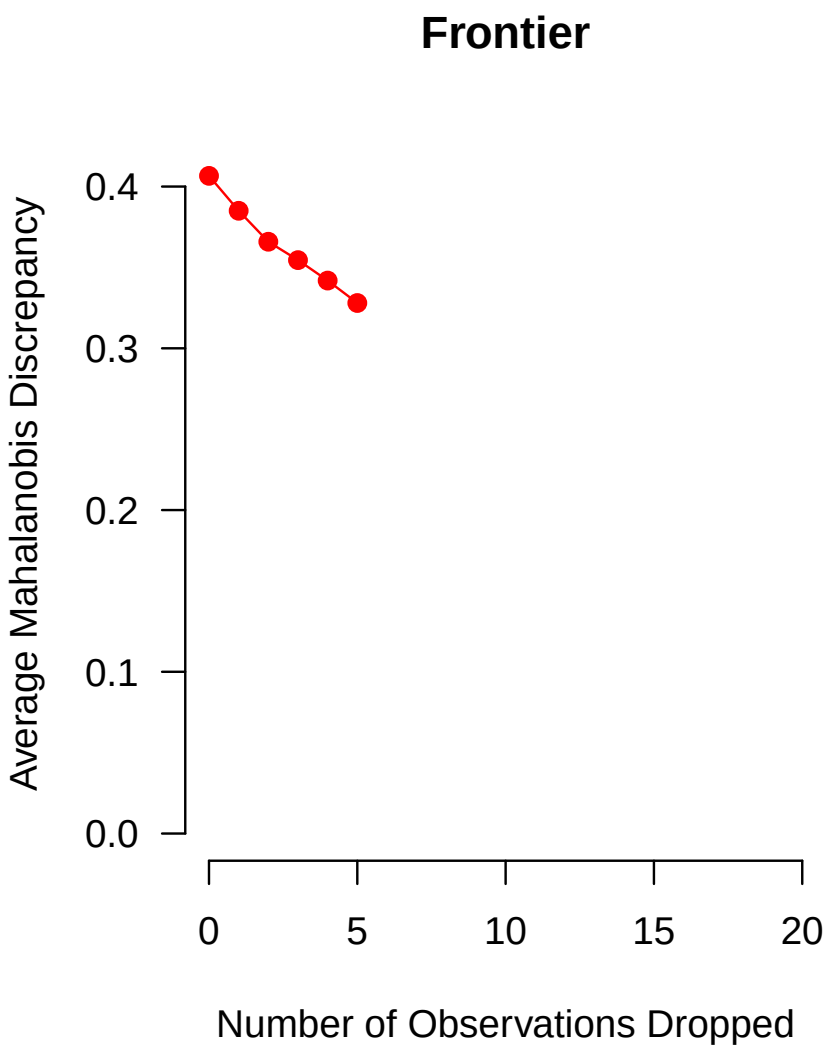
Constructing the FSATT Mahalanobis Frontier



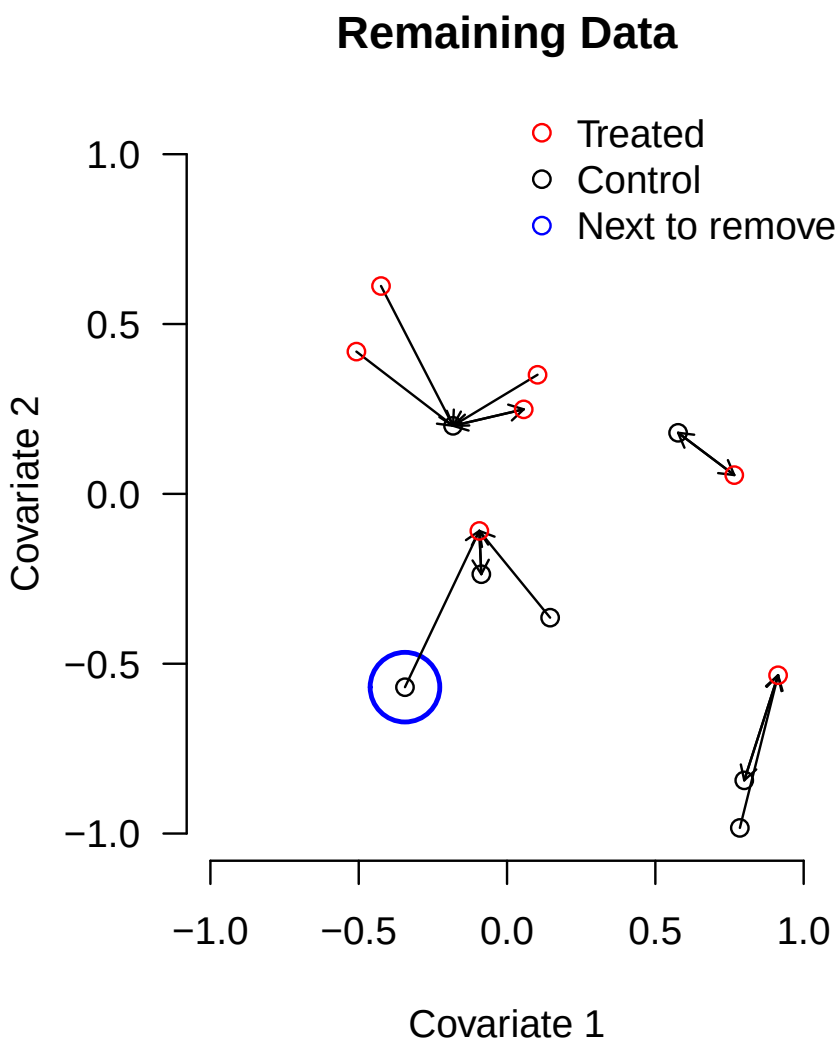


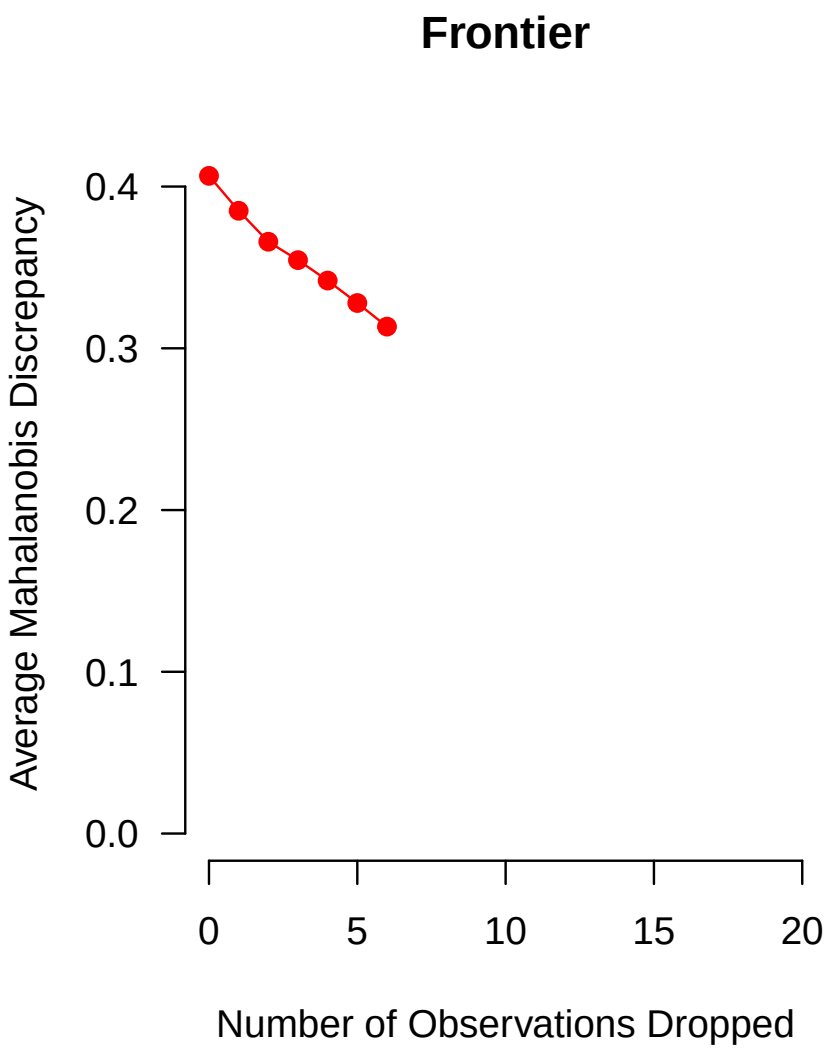
Constructing the FSATT Mahalanobis Frontier



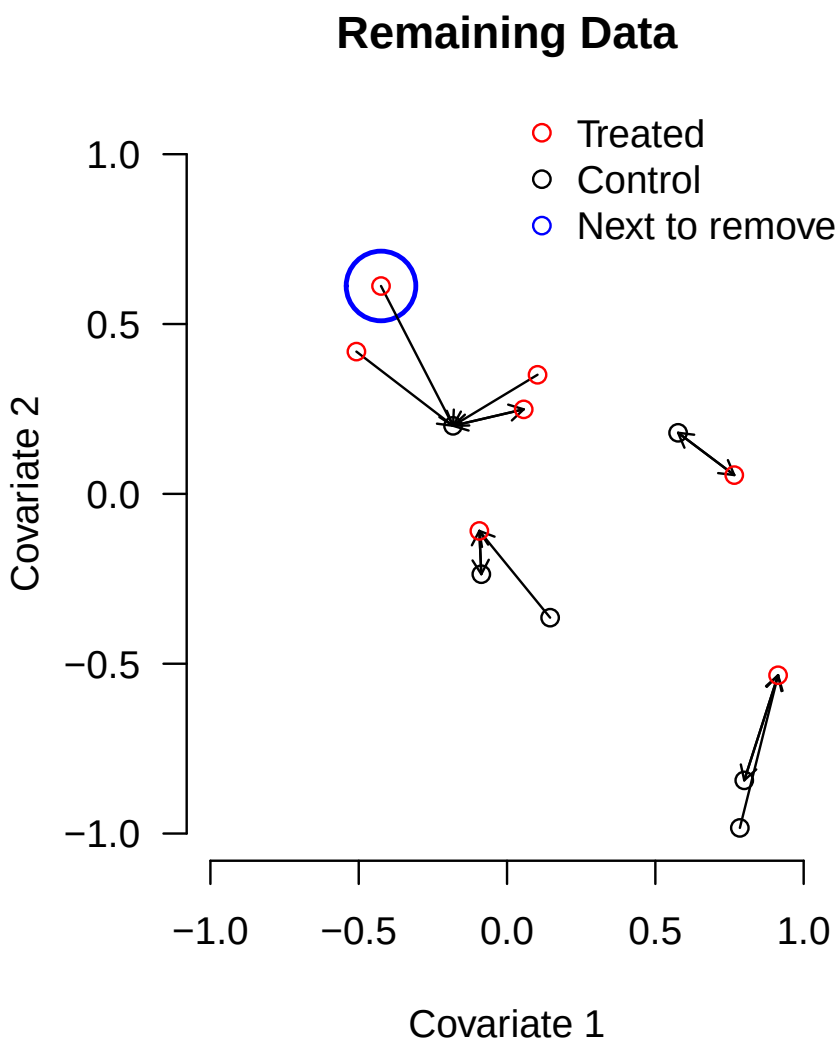


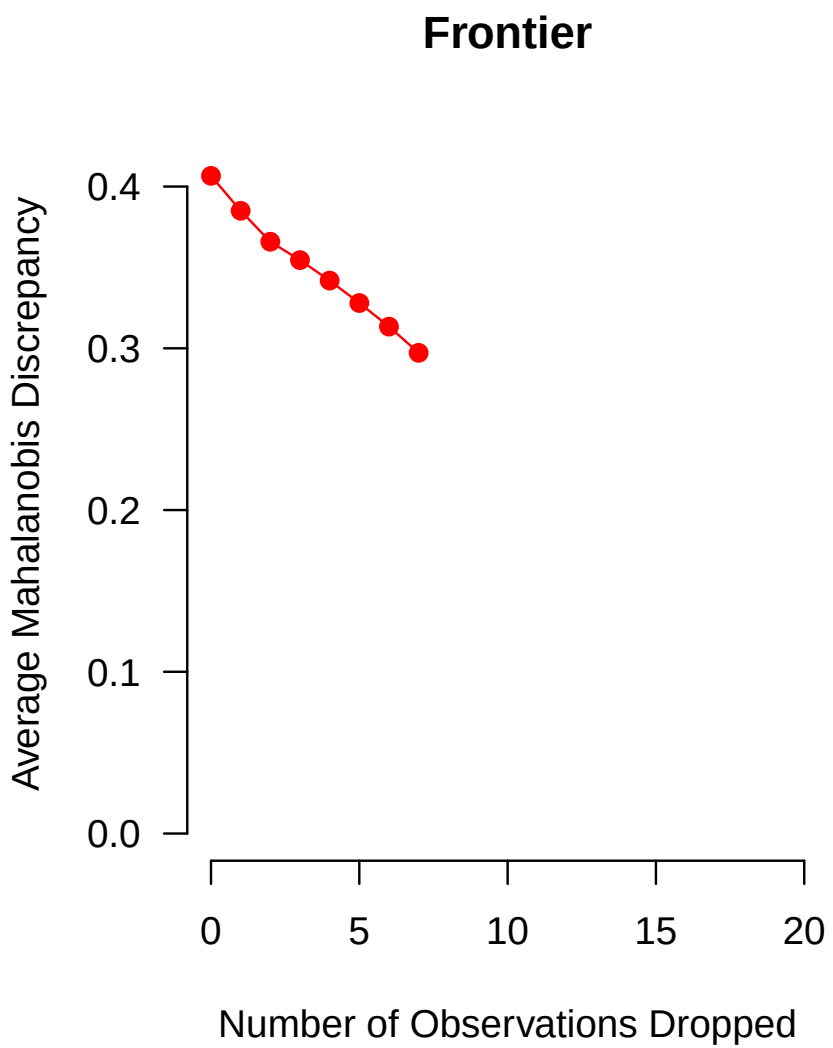
Constructing the FSATT Mahalanobis Frontier



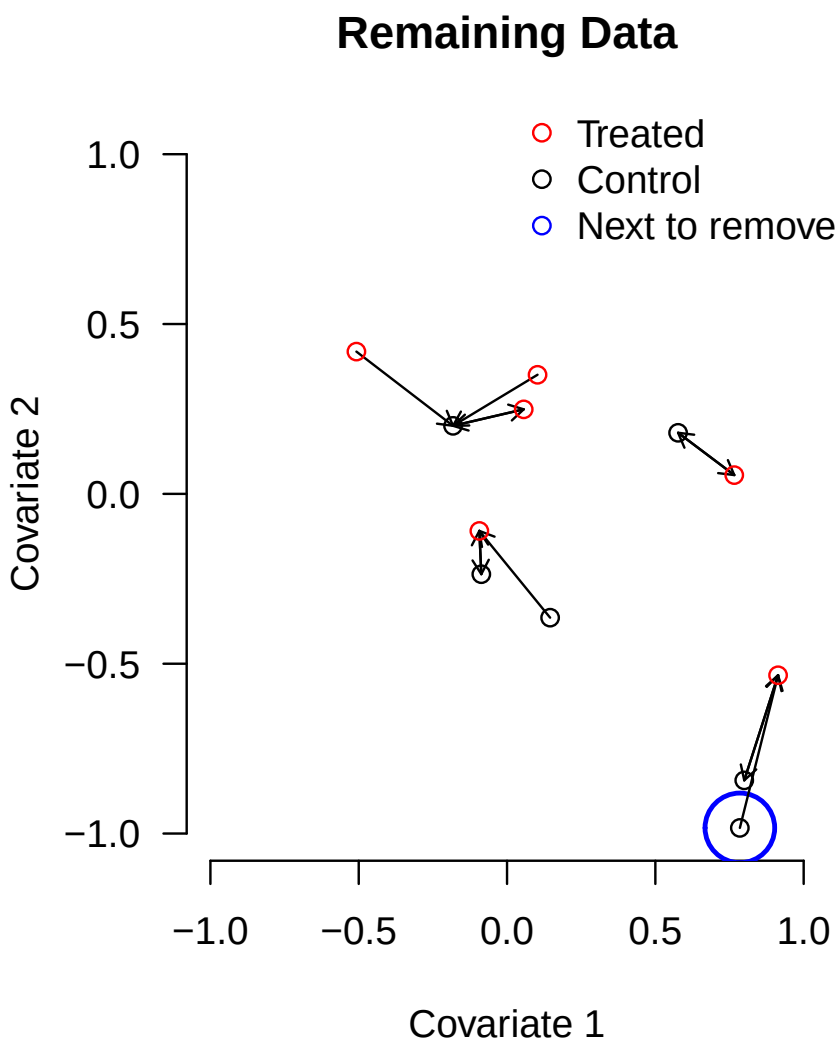


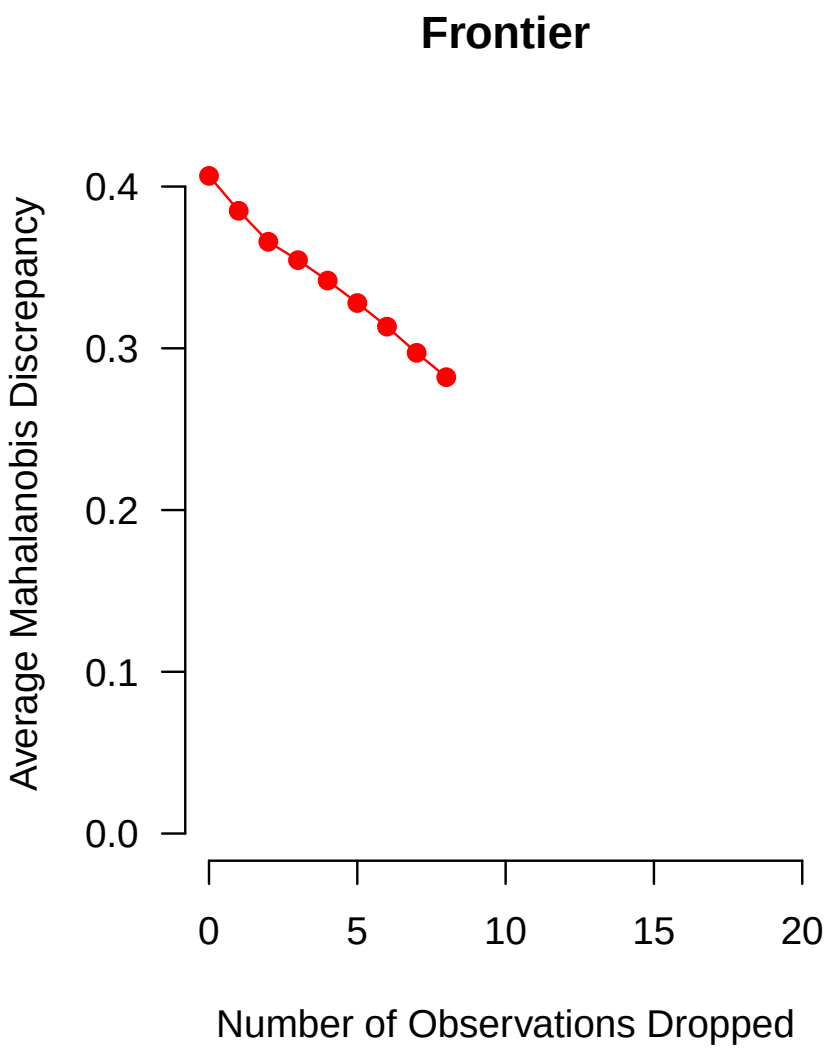
Constructing the FSATT Mahalanobis Frontier



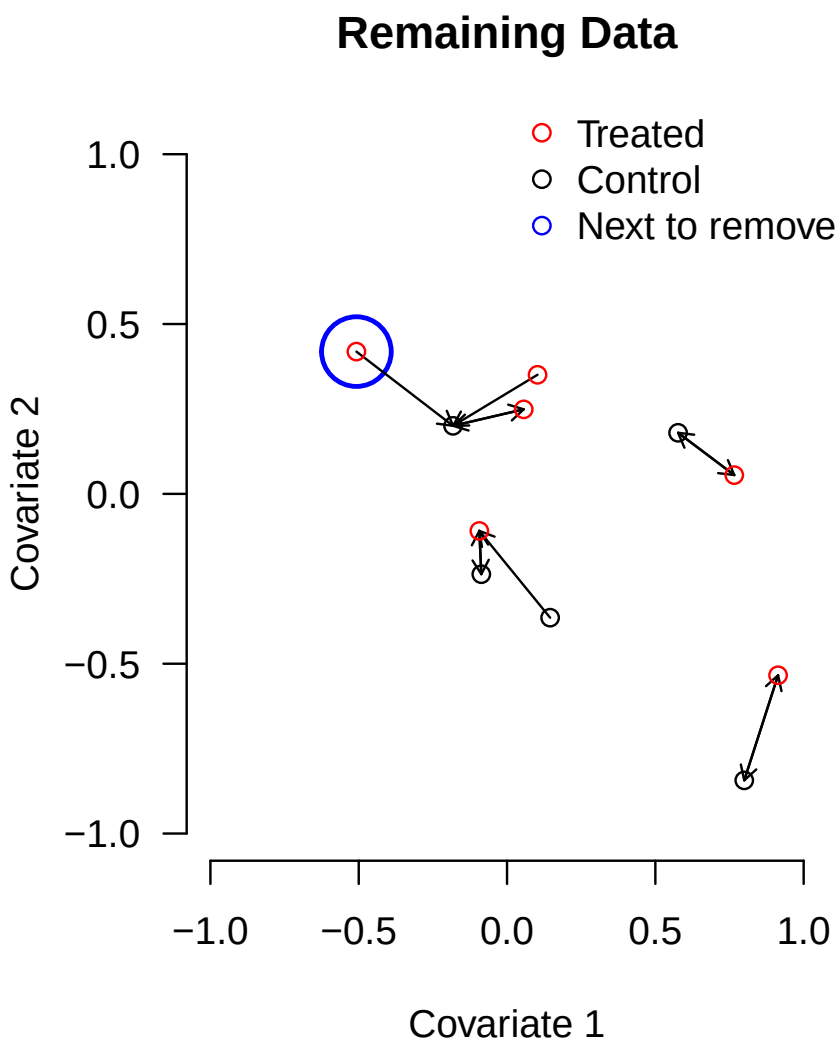


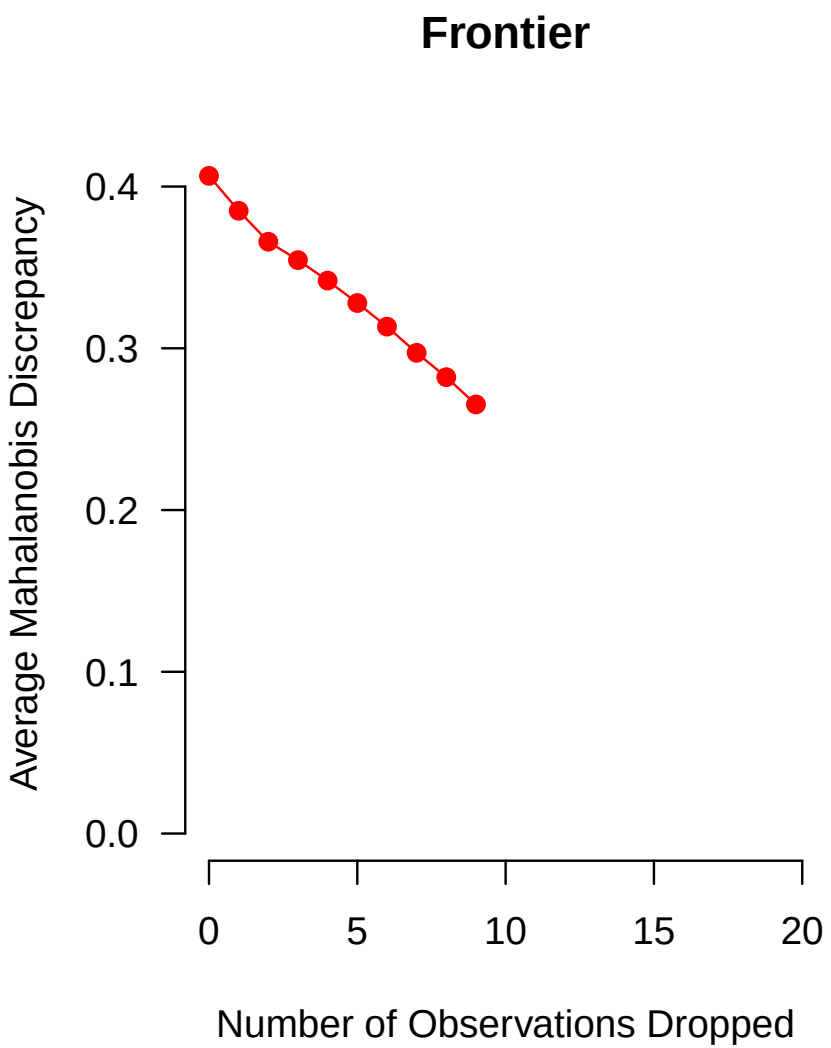
Constructing the FSATT Mahalanobis Frontier



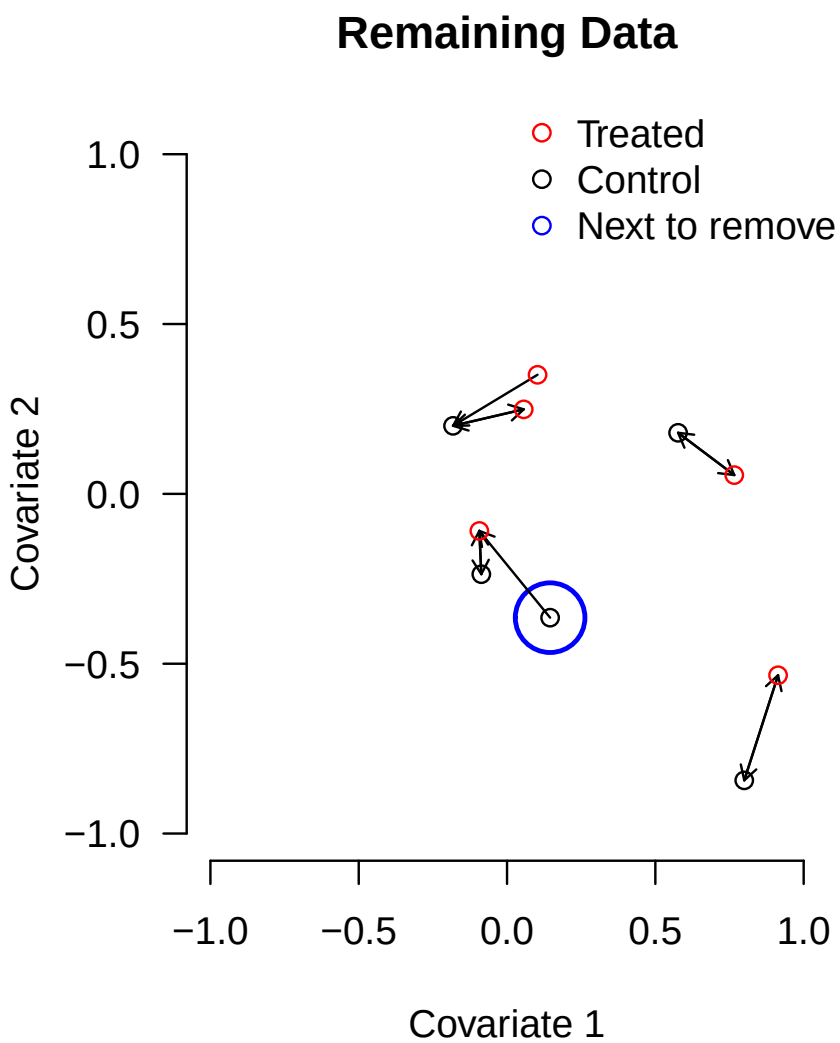


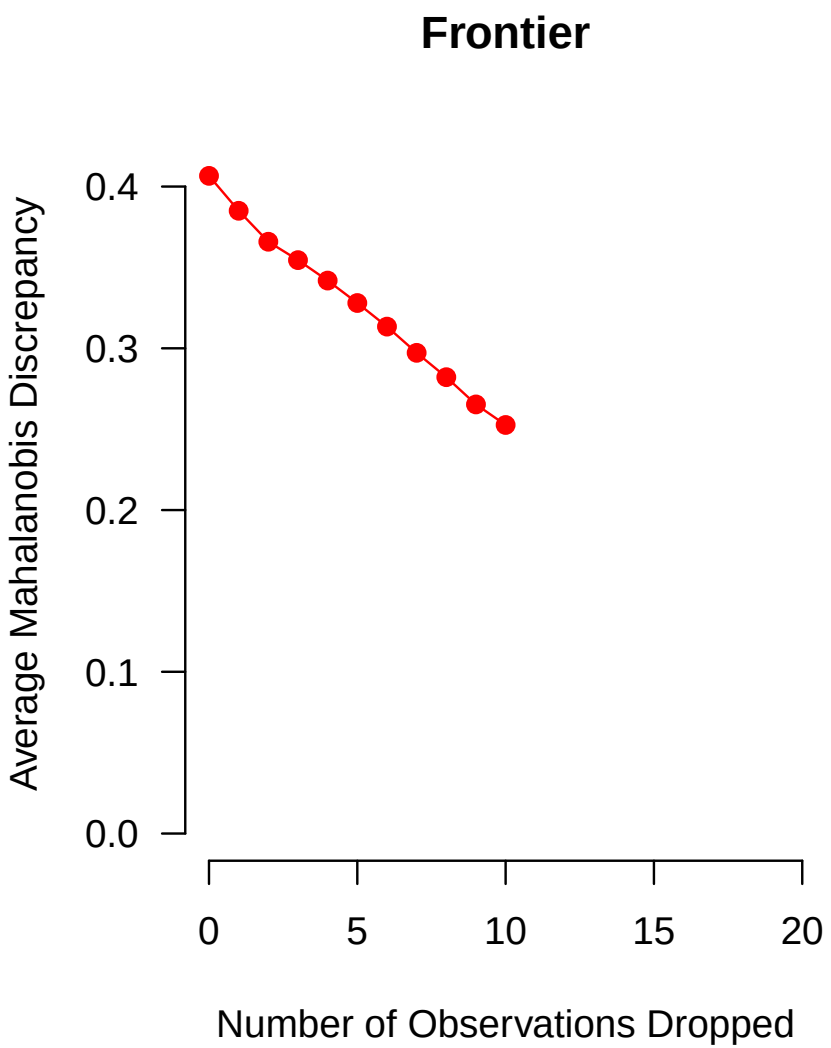
Constructing the FSATT Mahalanobis Frontier



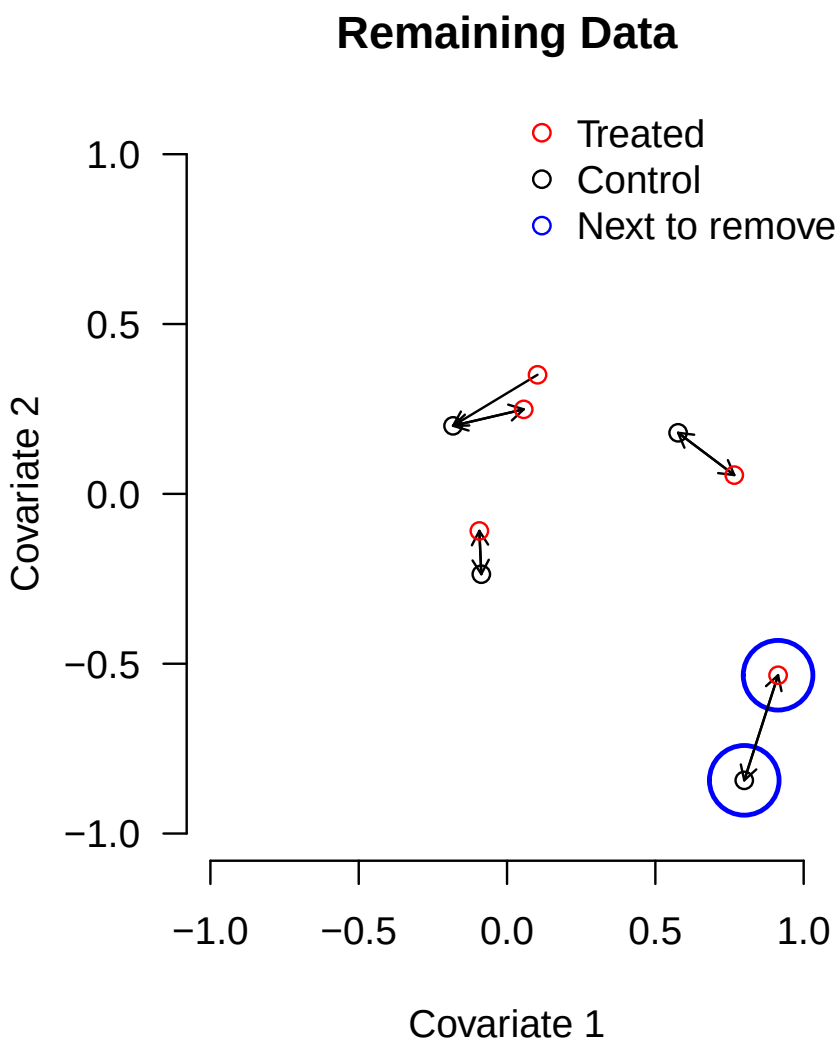


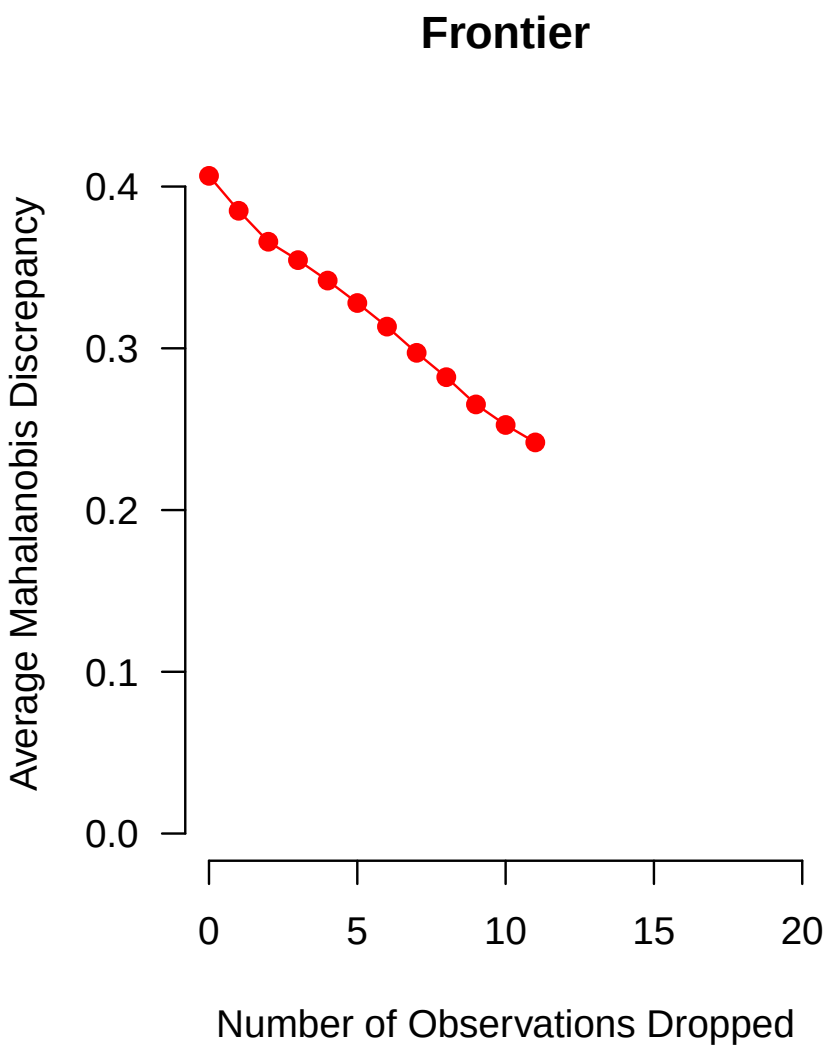
Constructing the FSATT Mahalanobis Frontier



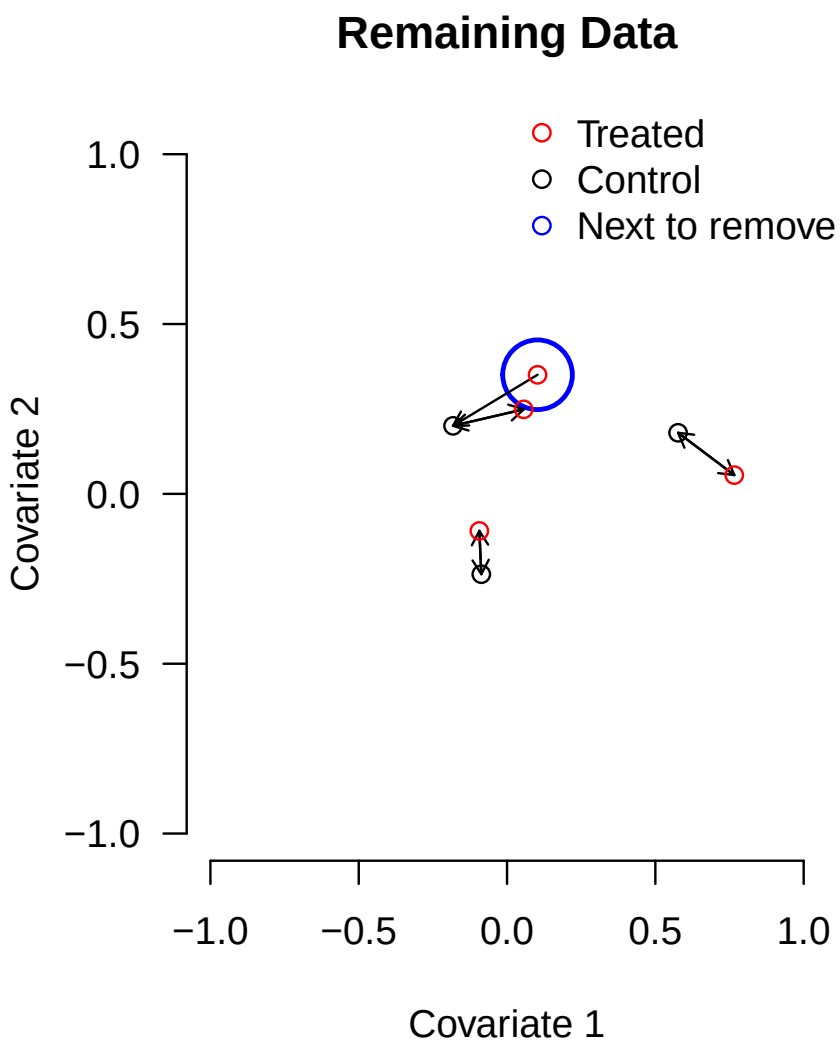


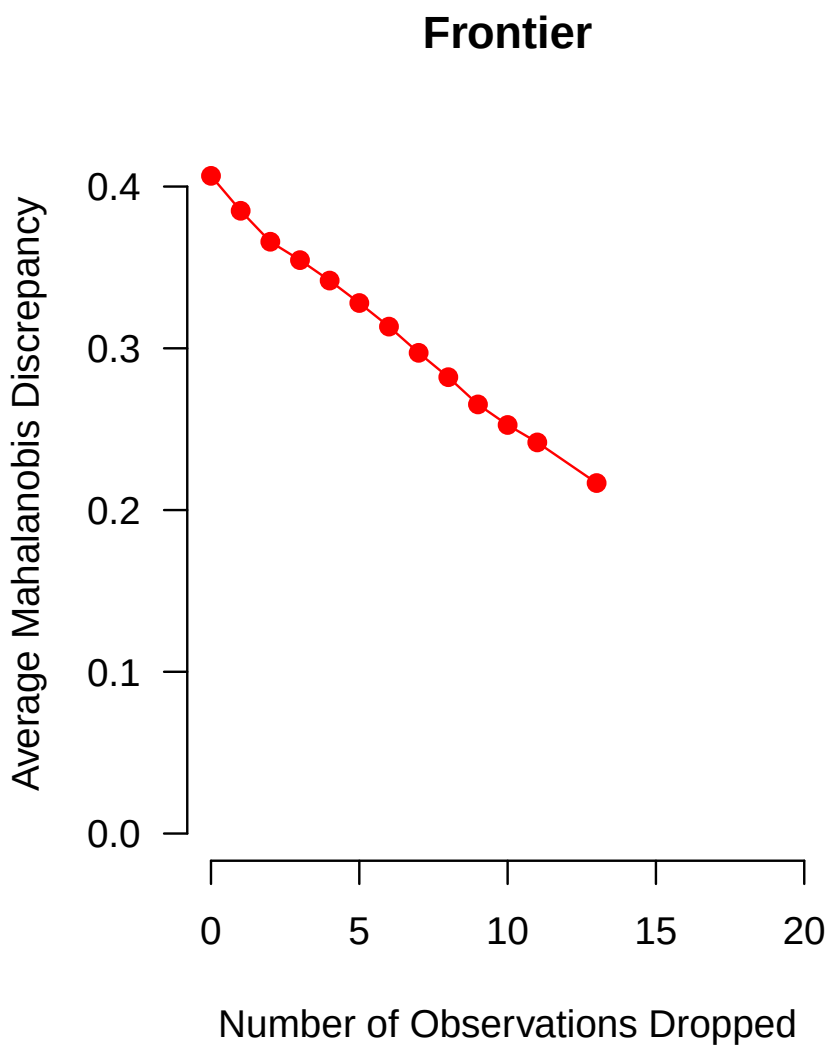
Constructing the FSATT Mahalanobis Frontier



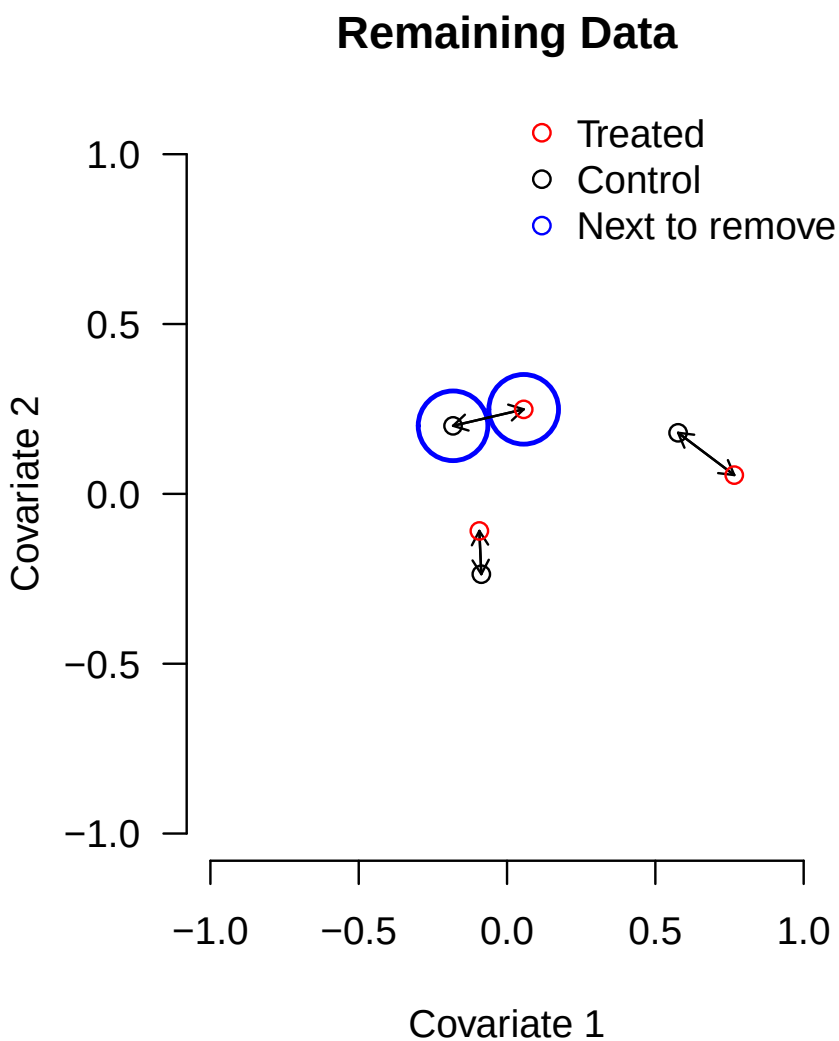


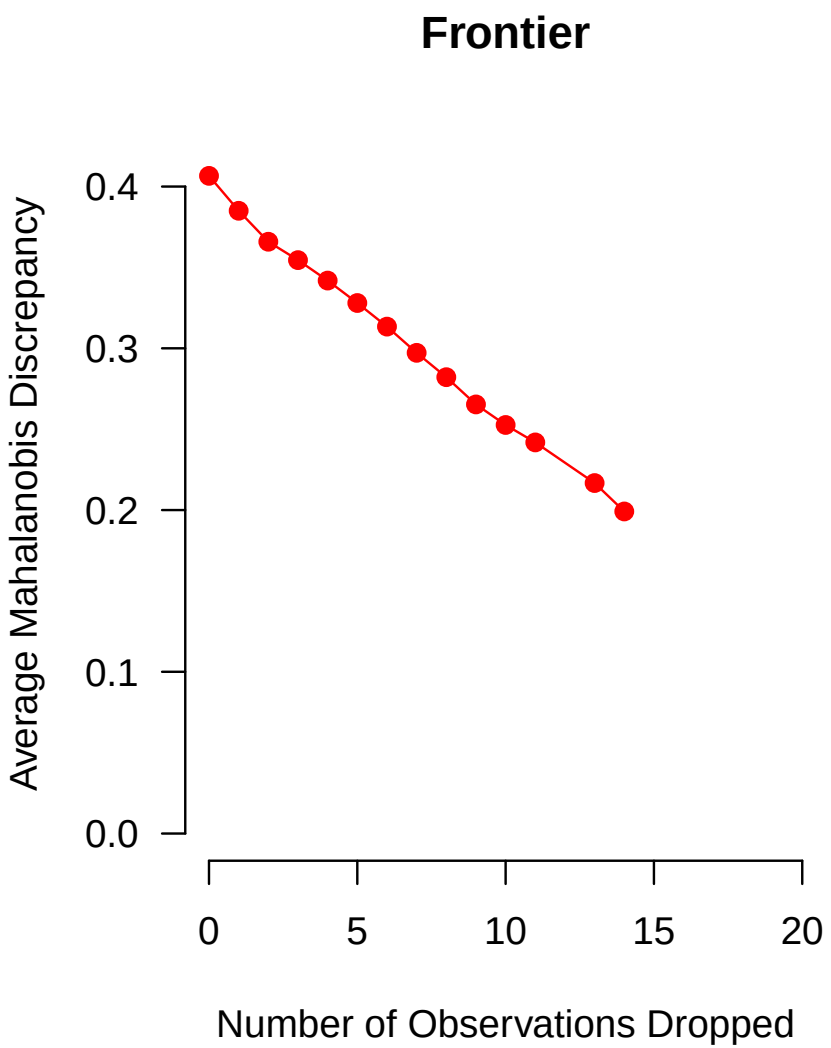
Constructing the FSATT Mahalanobis Frontier



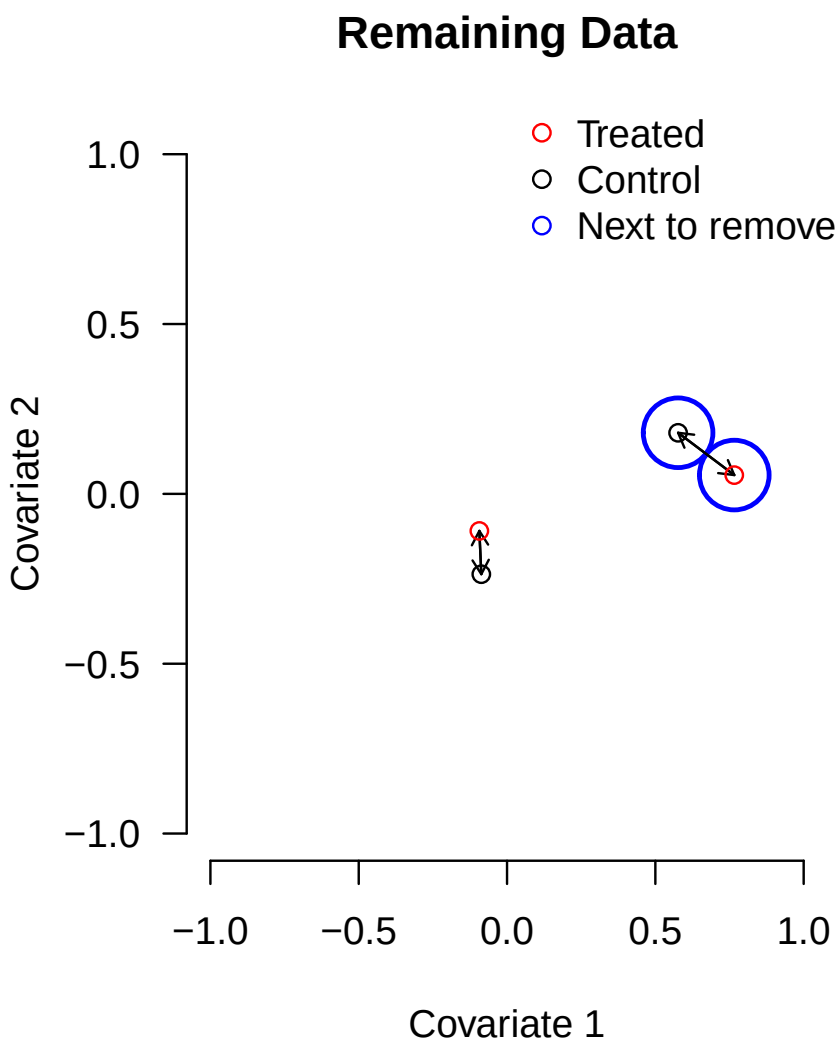


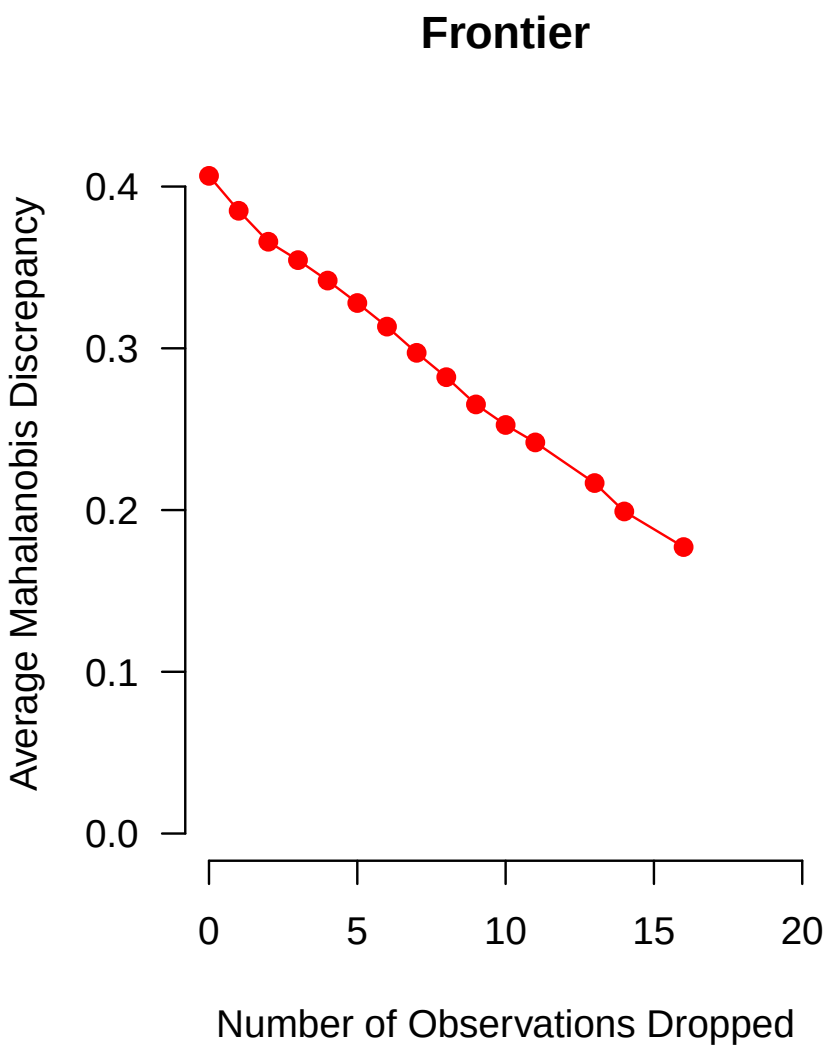
Constructing the FSATT Mahalanobis Frontier



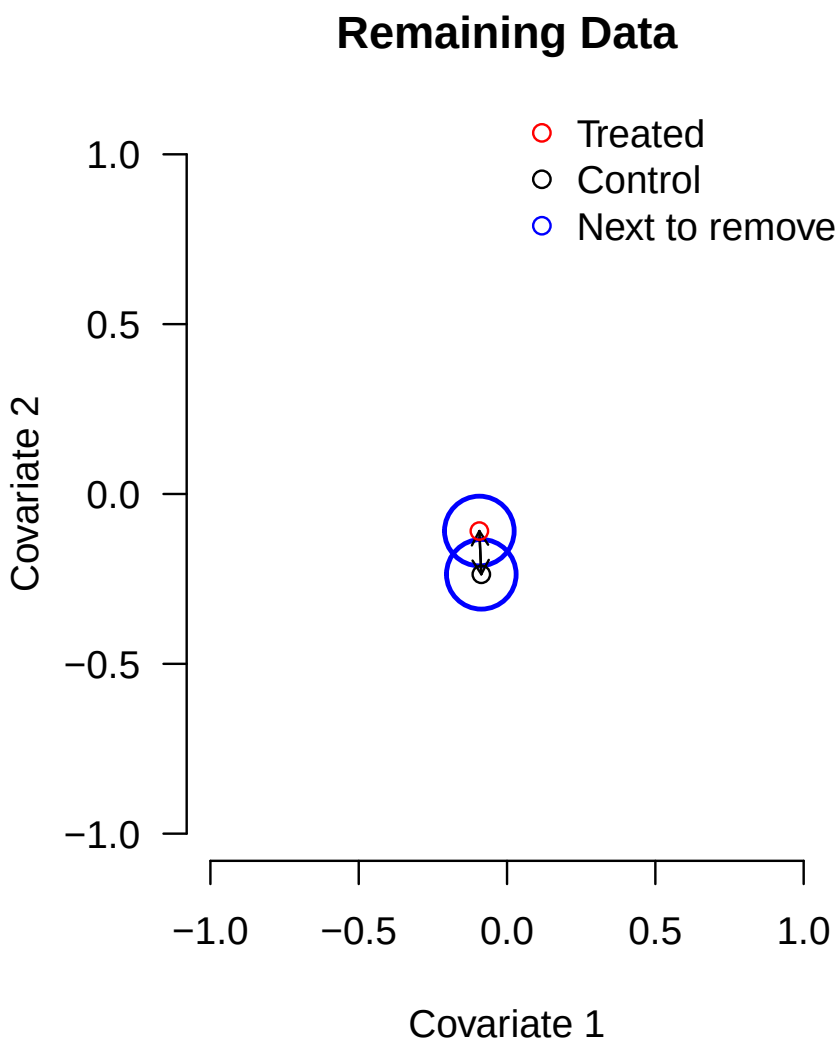


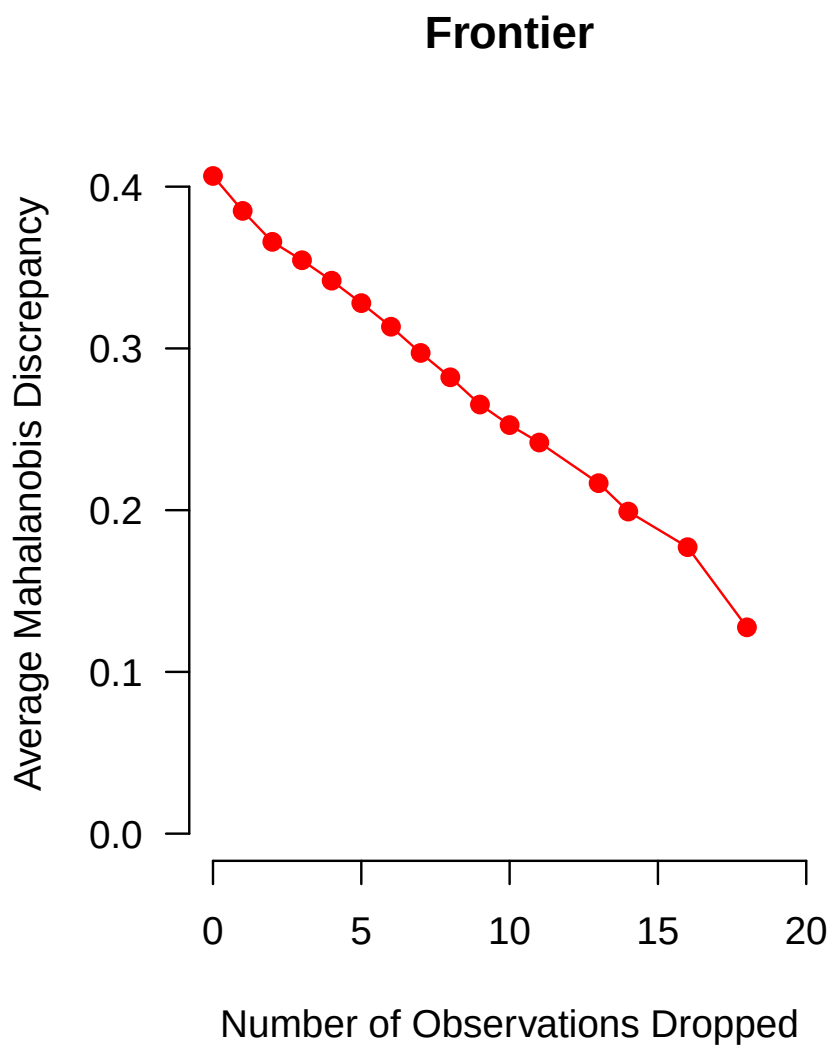
Constructing the FSATT Mahalanobis Frontier



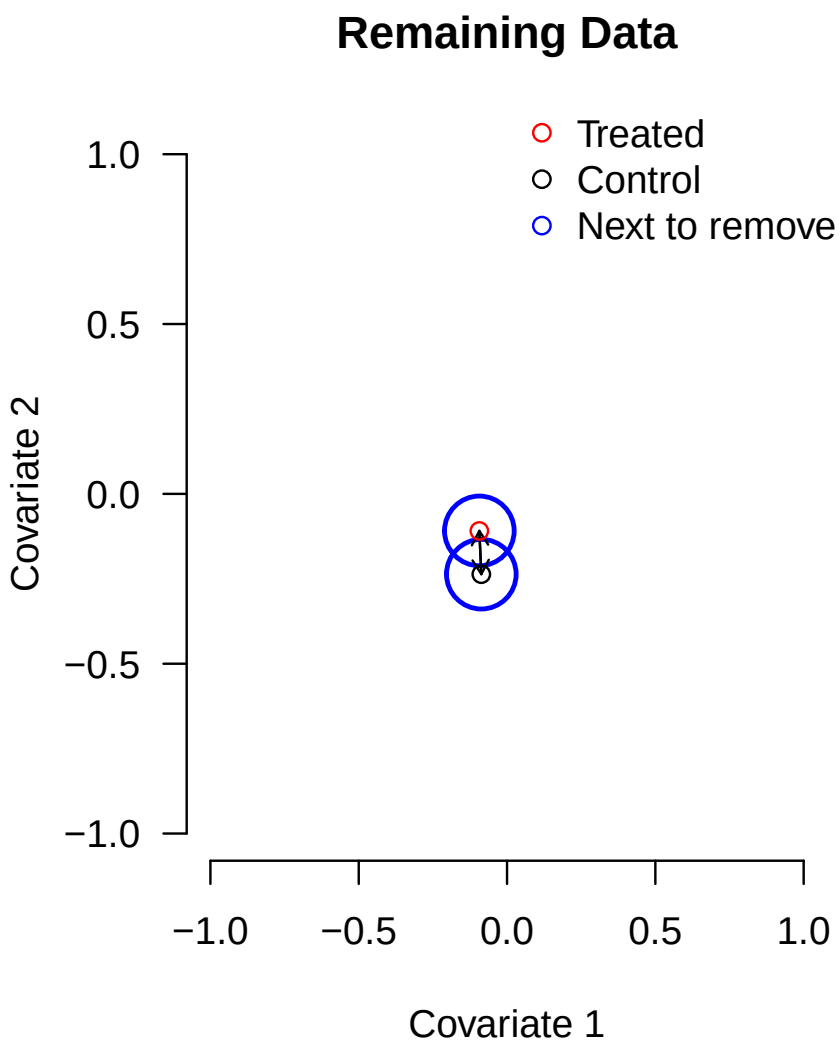


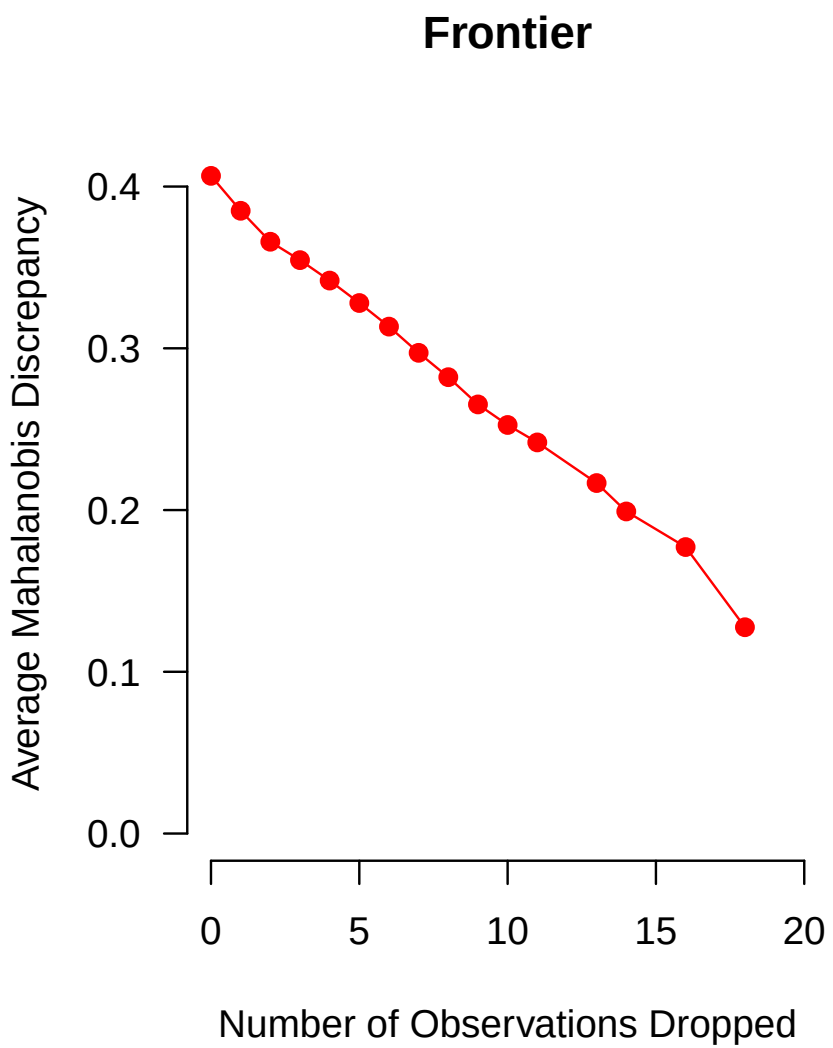
Constructing the FSATT Mahalanobis Frontier





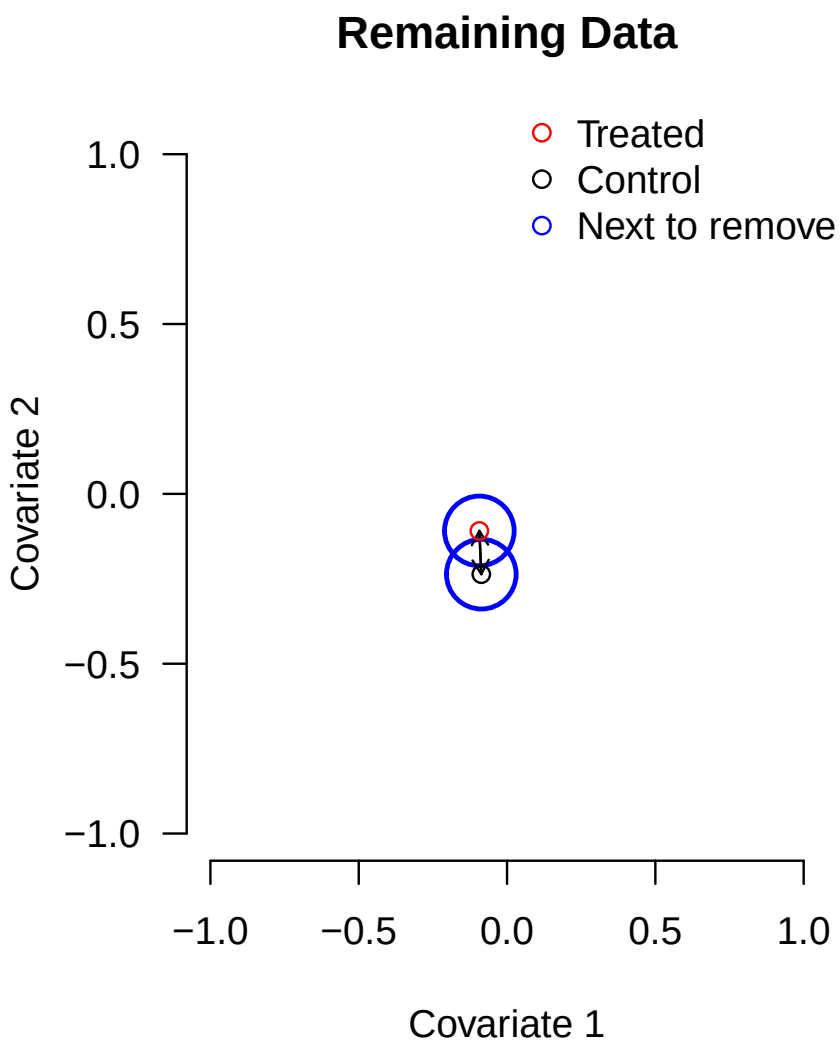
Constructing the FSATT Mahalanobis Frontier

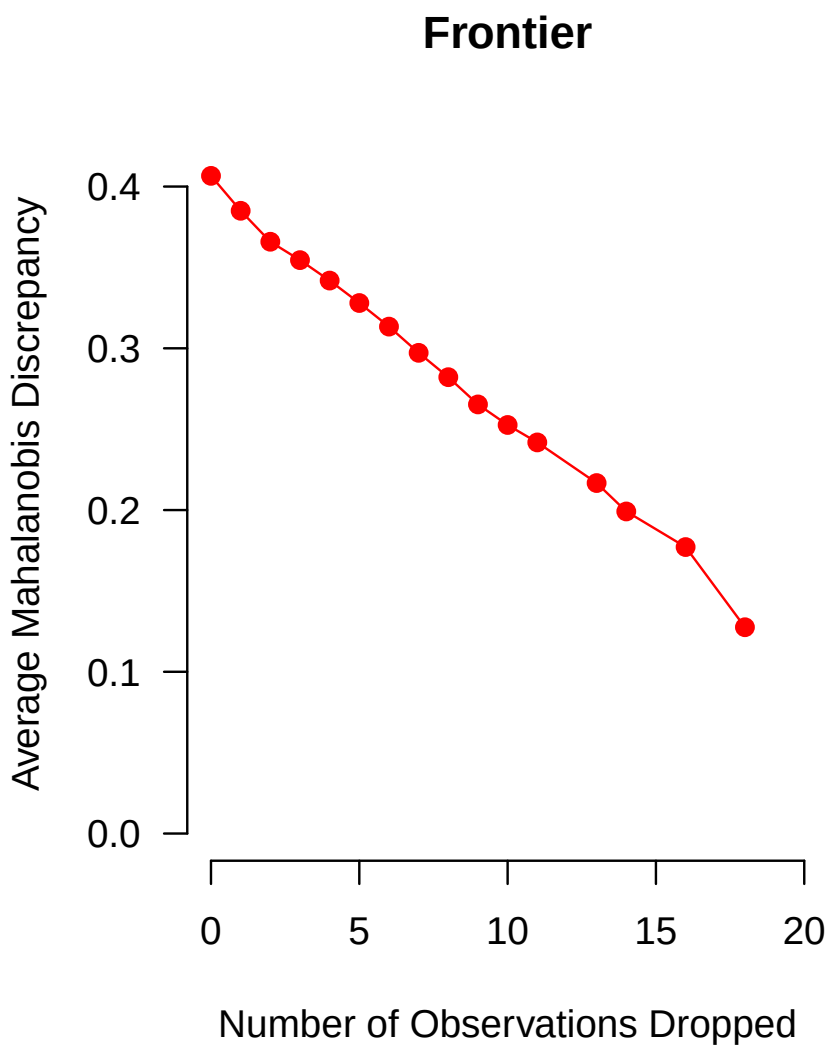




- Warning: figure omits some details!

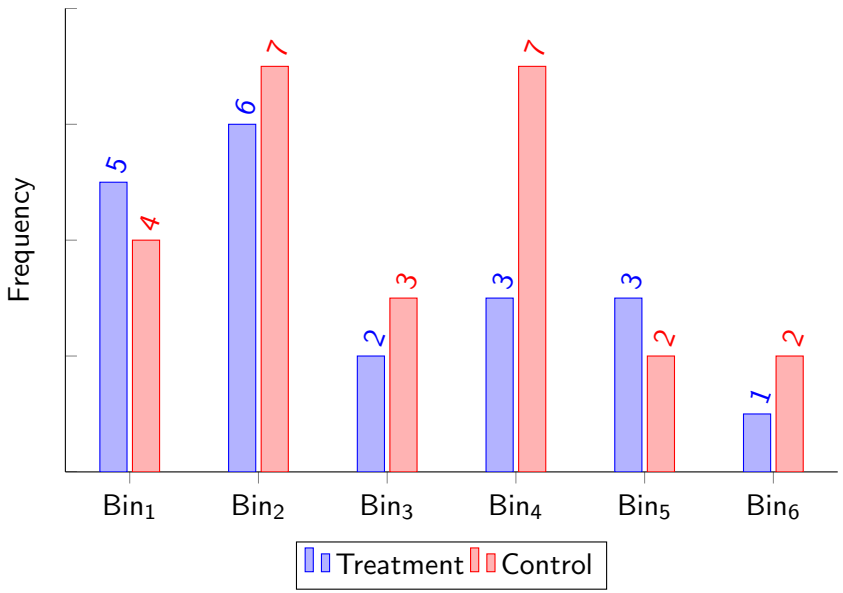
Constructing the FSATT Mahalanobis Frontier

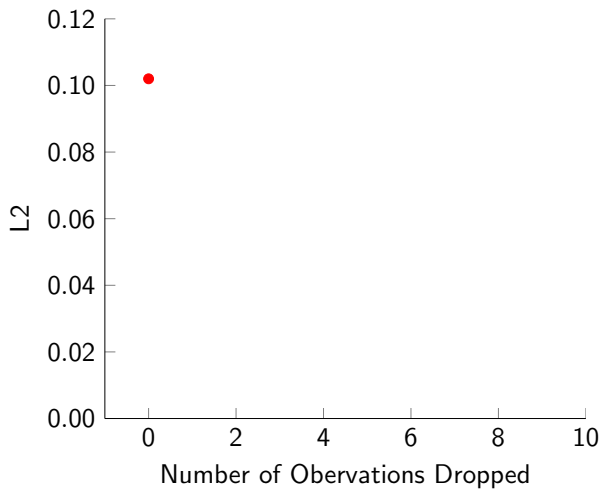




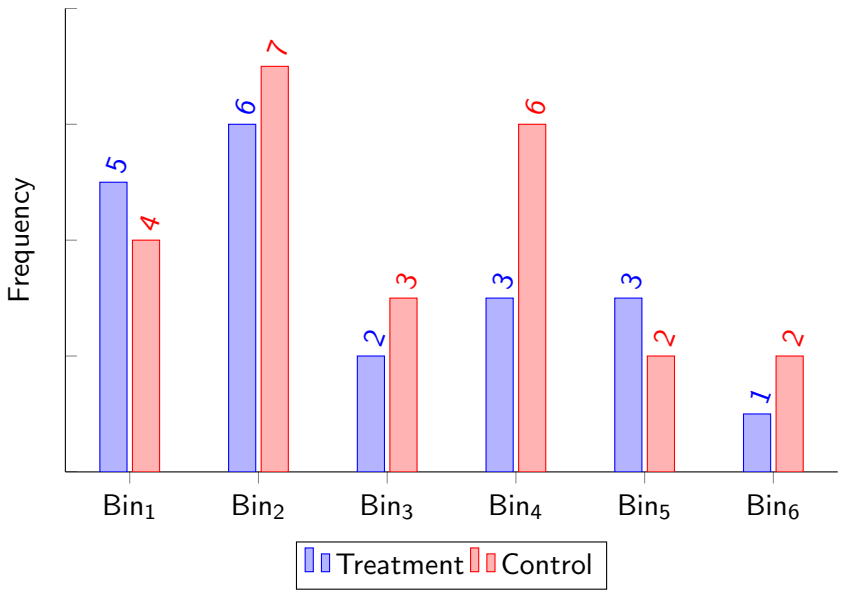
- Warning: figure omits some details!
- Very fast; works with any continuous imbalance metric

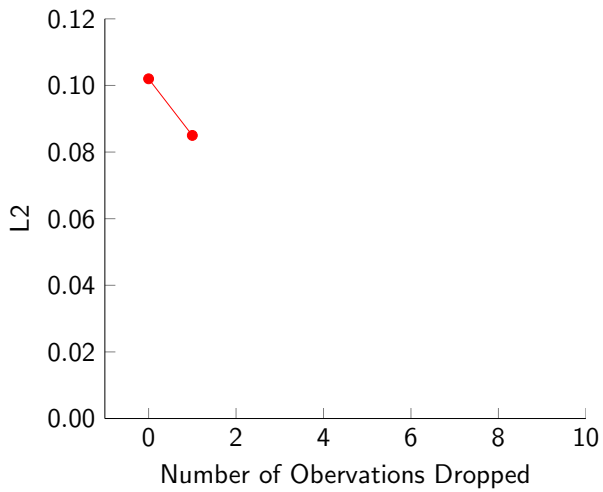
Constructing the L1/L2 SATT Frontier



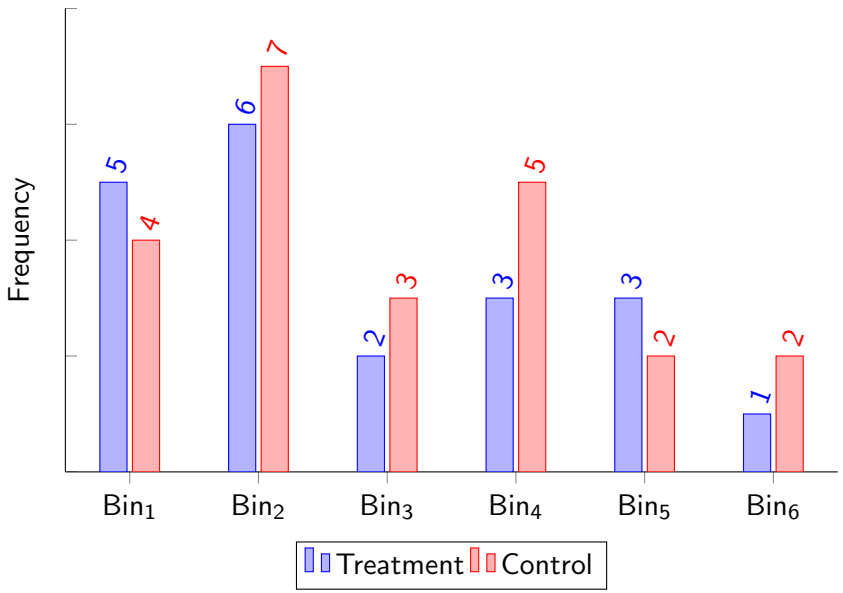


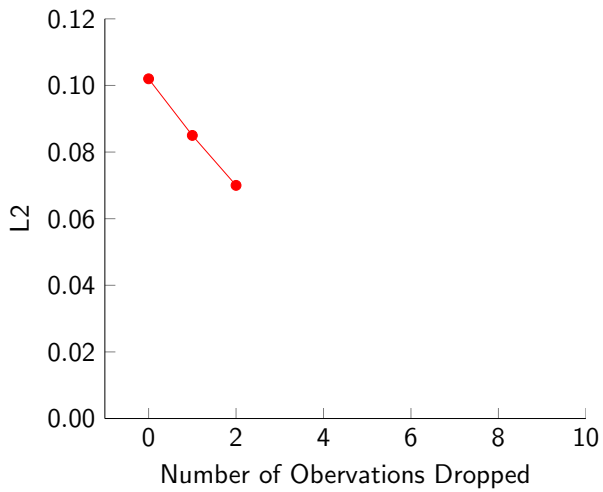
Constructing the L1/L2 SATT Frontier



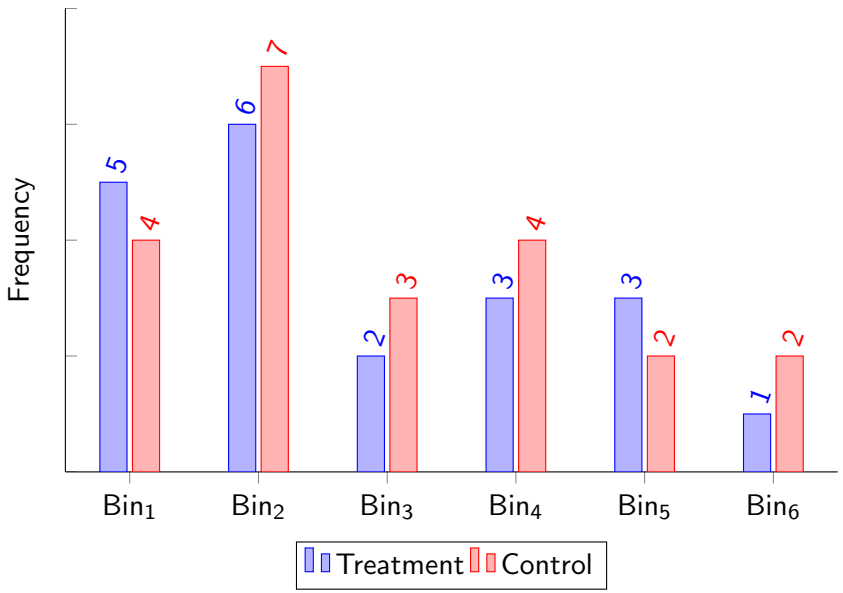


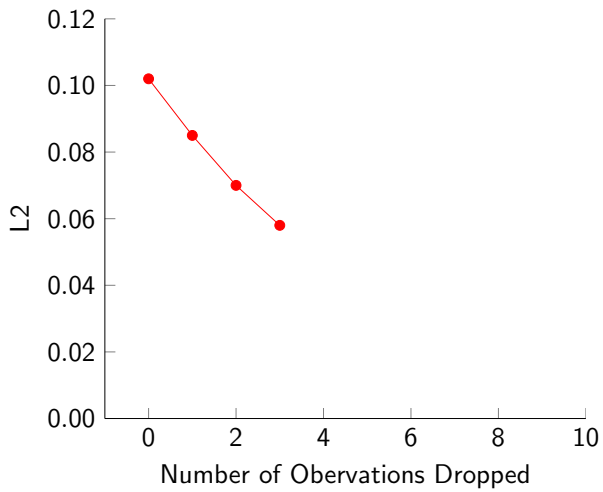
Constructing the L1/L2 SATT Frontier



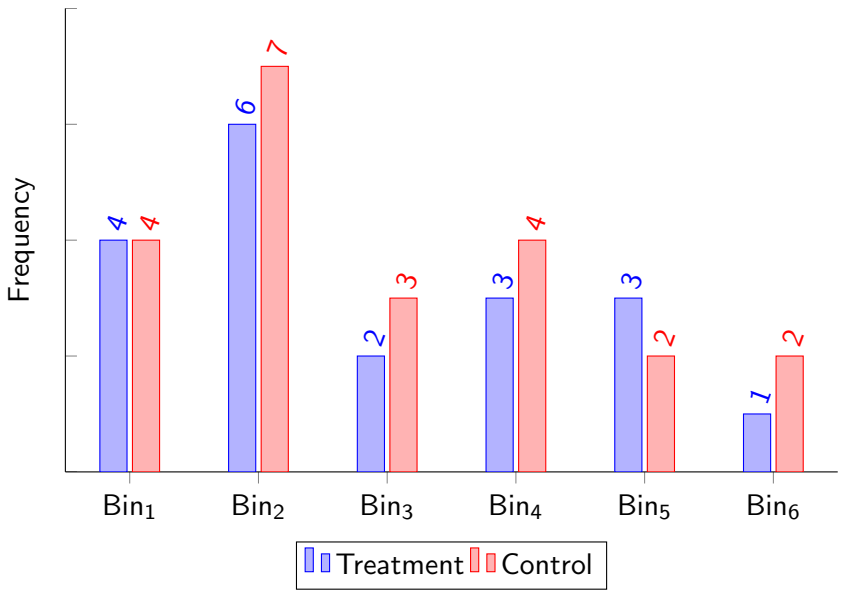


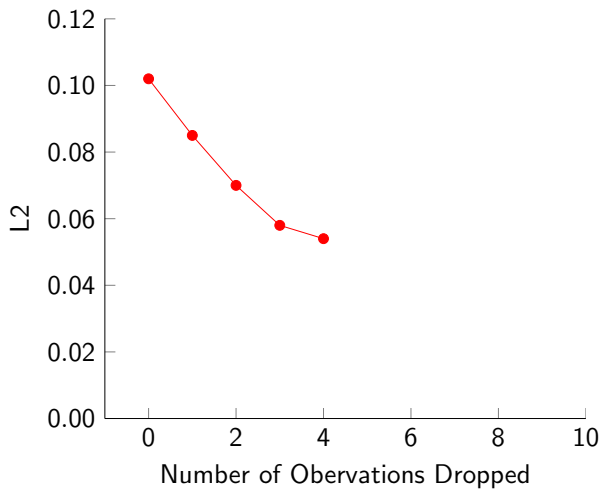
Constructing the L1/L2 SATT Frontier



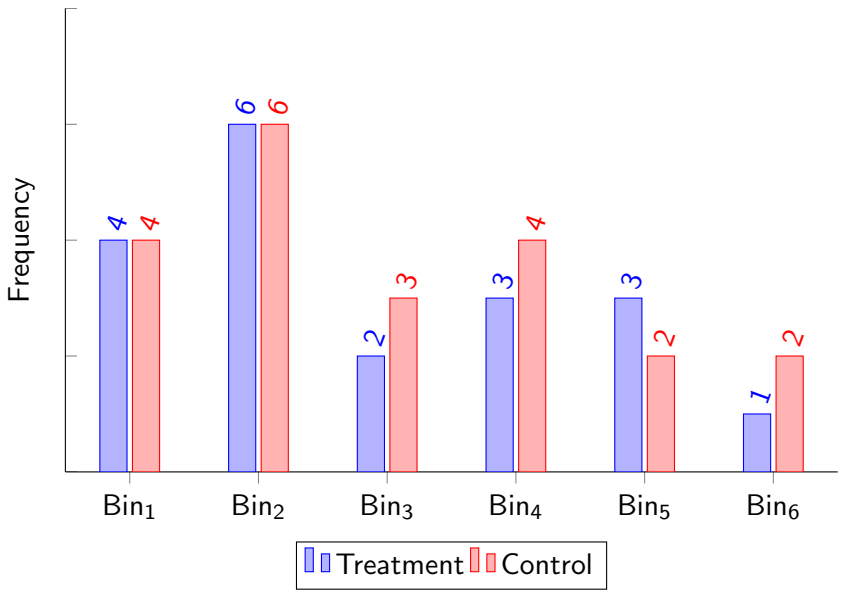


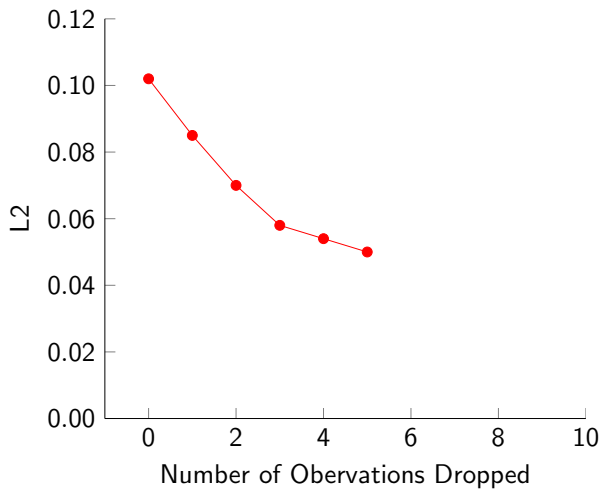
Constructing the L1/L2 SATT Frontier



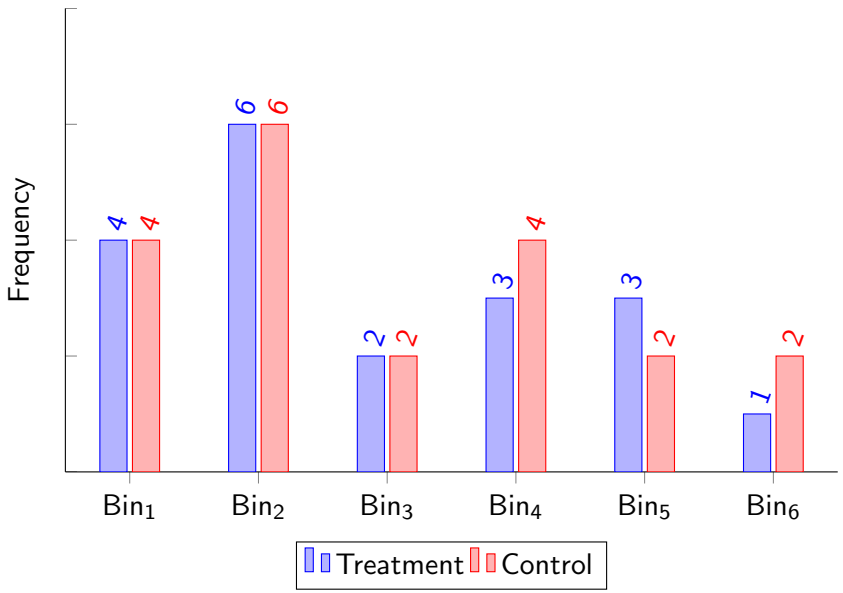


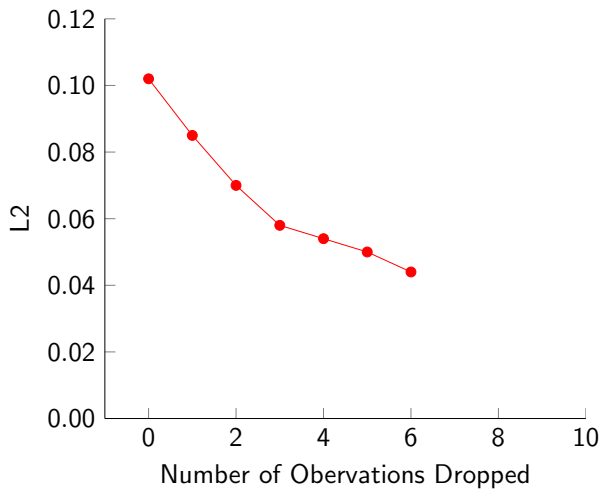
Constructing the L1/L2 SATT Frontier



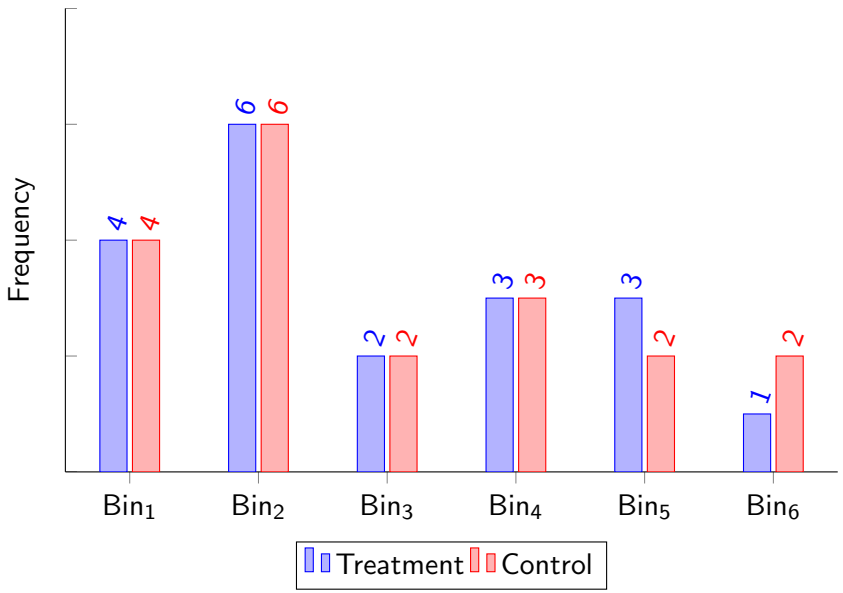


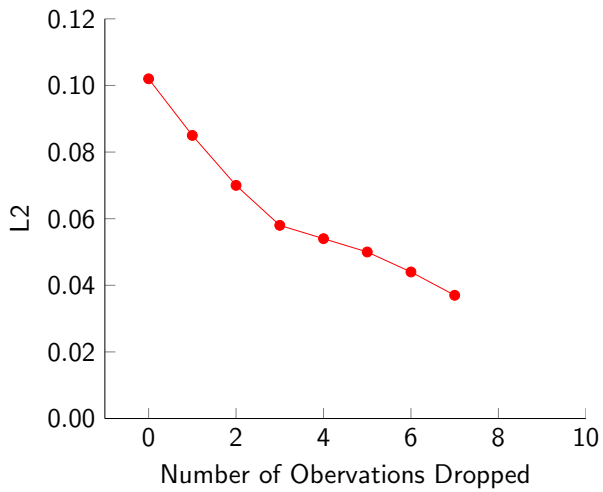
Constructing the L1/L2 SATT Frontier



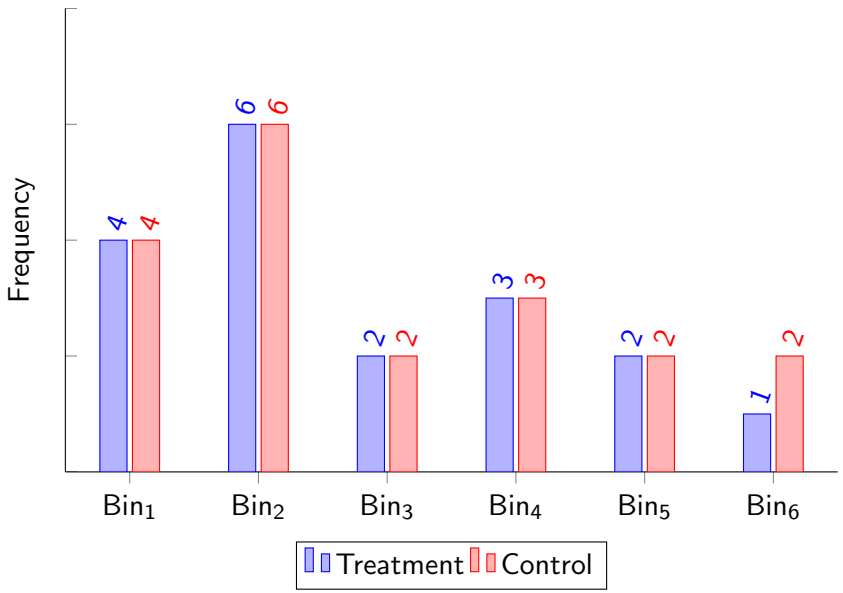


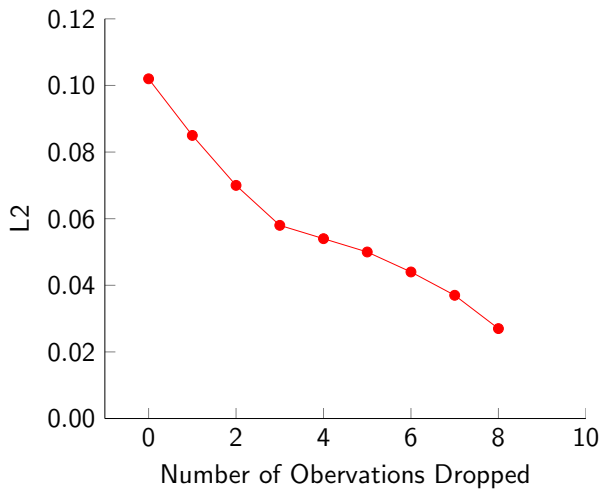
Constructing the L1/L2 SATT Frontier



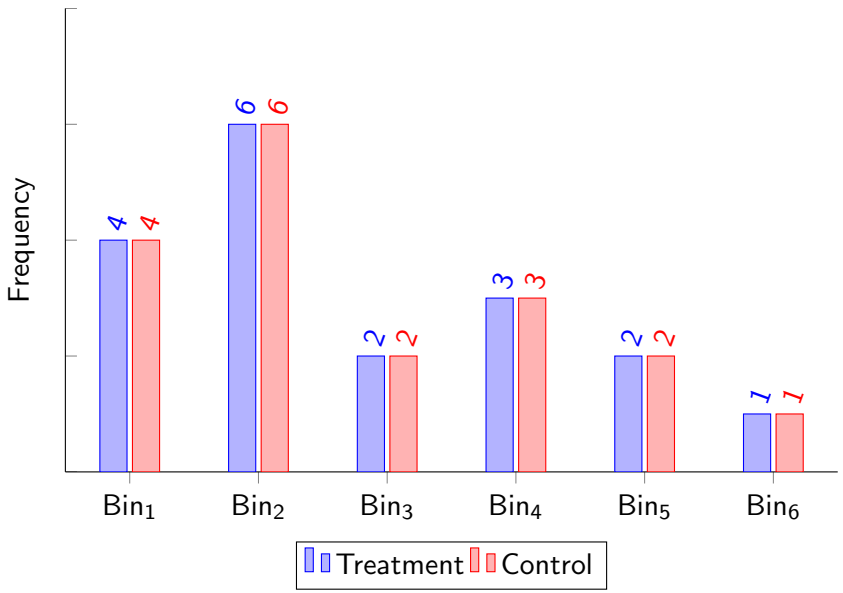


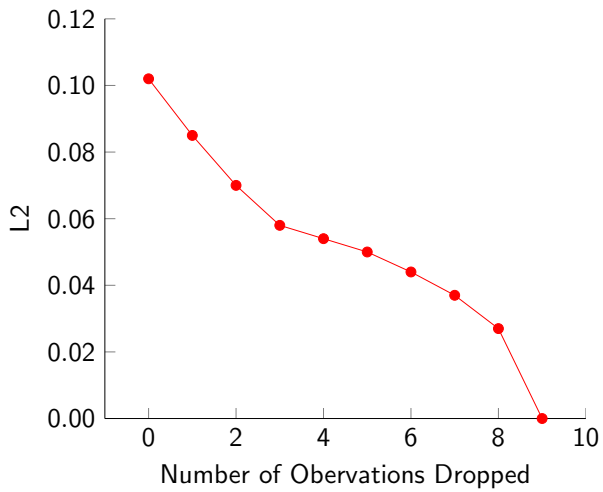
Constructing the L1/L2 SATT Frontier





Constructing the L1/L2 SATT Frontier





Constructing the L1/L2 SATT Frontier

Constructing the L1/L2 SATT Frontier

- Warning: This figure omits some technical details too!

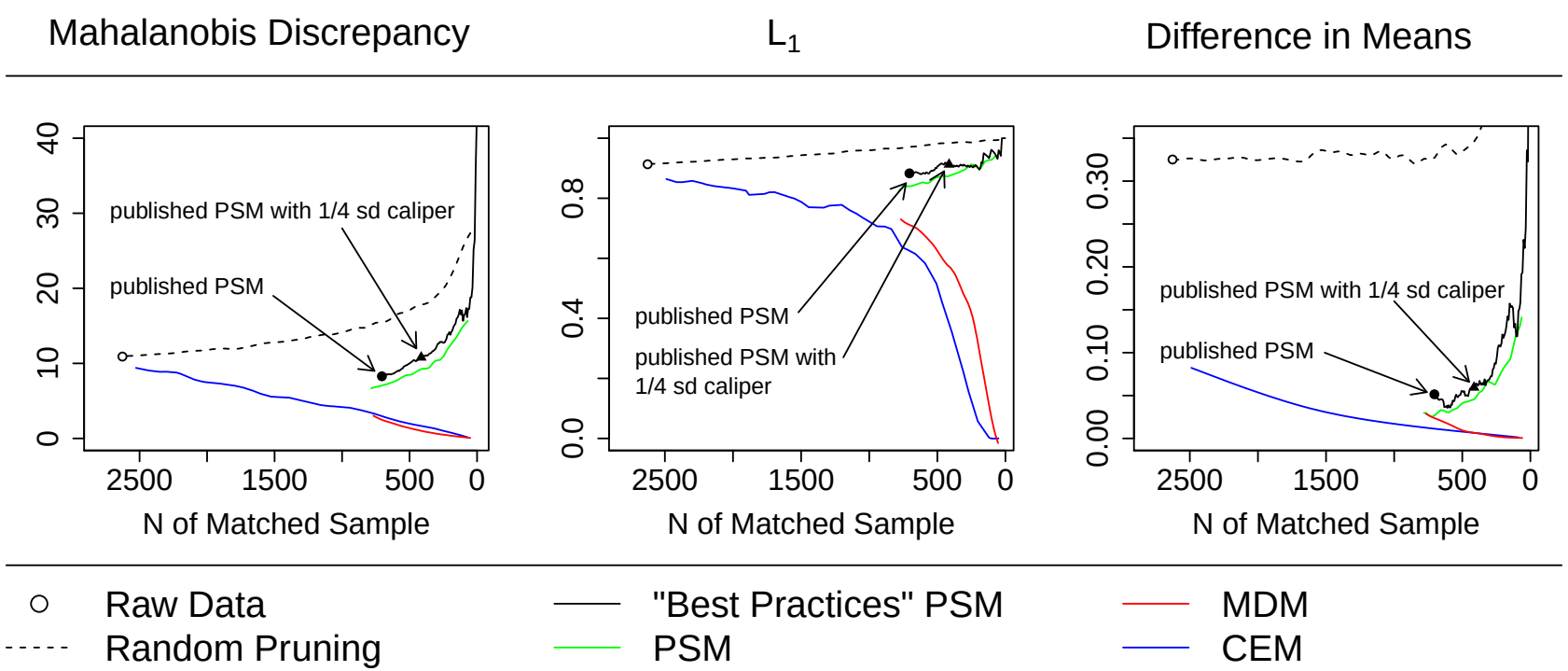
Constructing the L1/L2 SATT Frontier

- Warning: This figure omits some technical details too!
- Works very fast, even with very large data sets

Problems with PSM: Foreign Aid Shocks

King, Nielsen, Coberley, Pope, and Wells (2012)

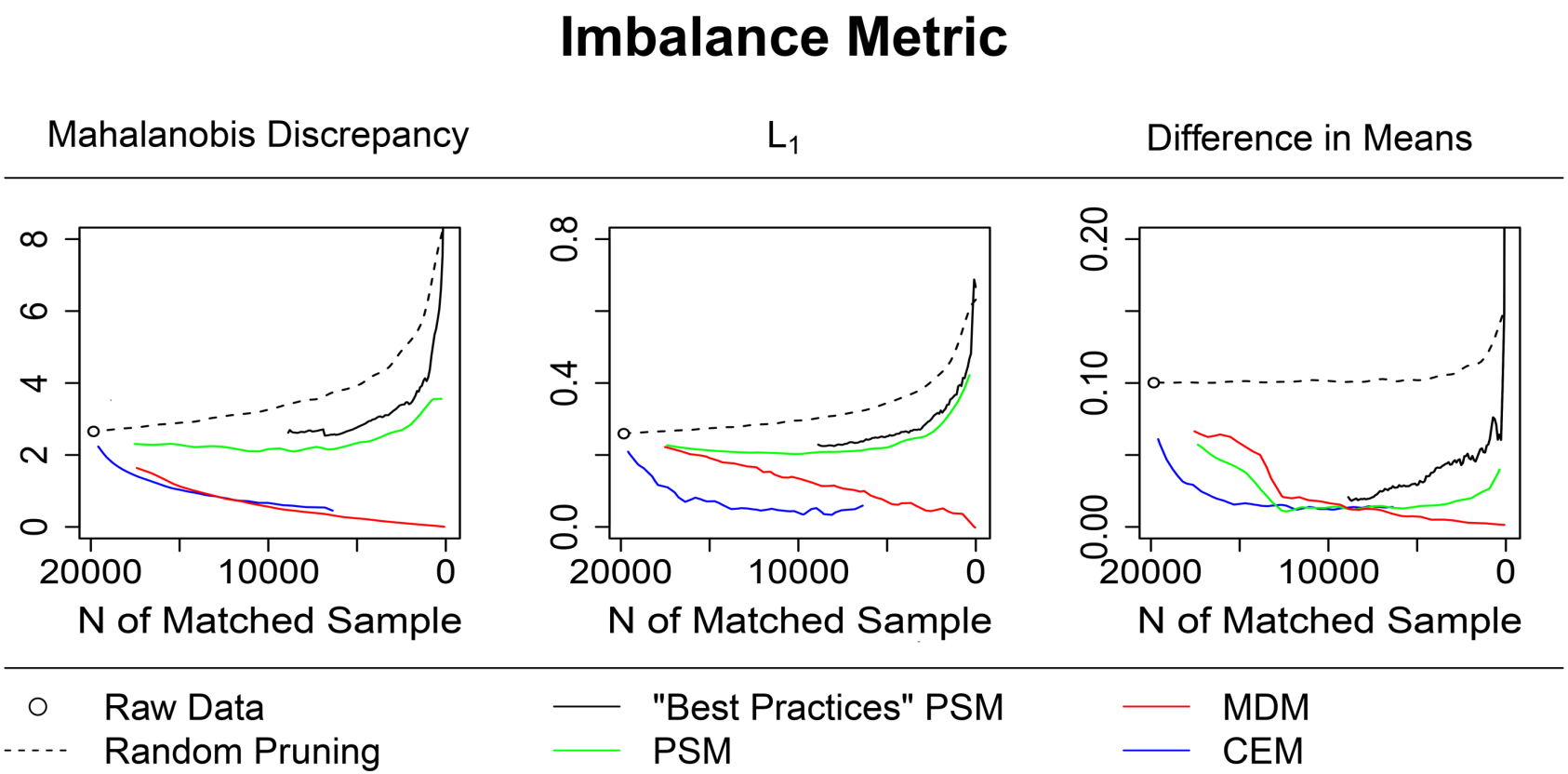
Imbalance Metric



Methods-specific frontiers (for methodological research only)

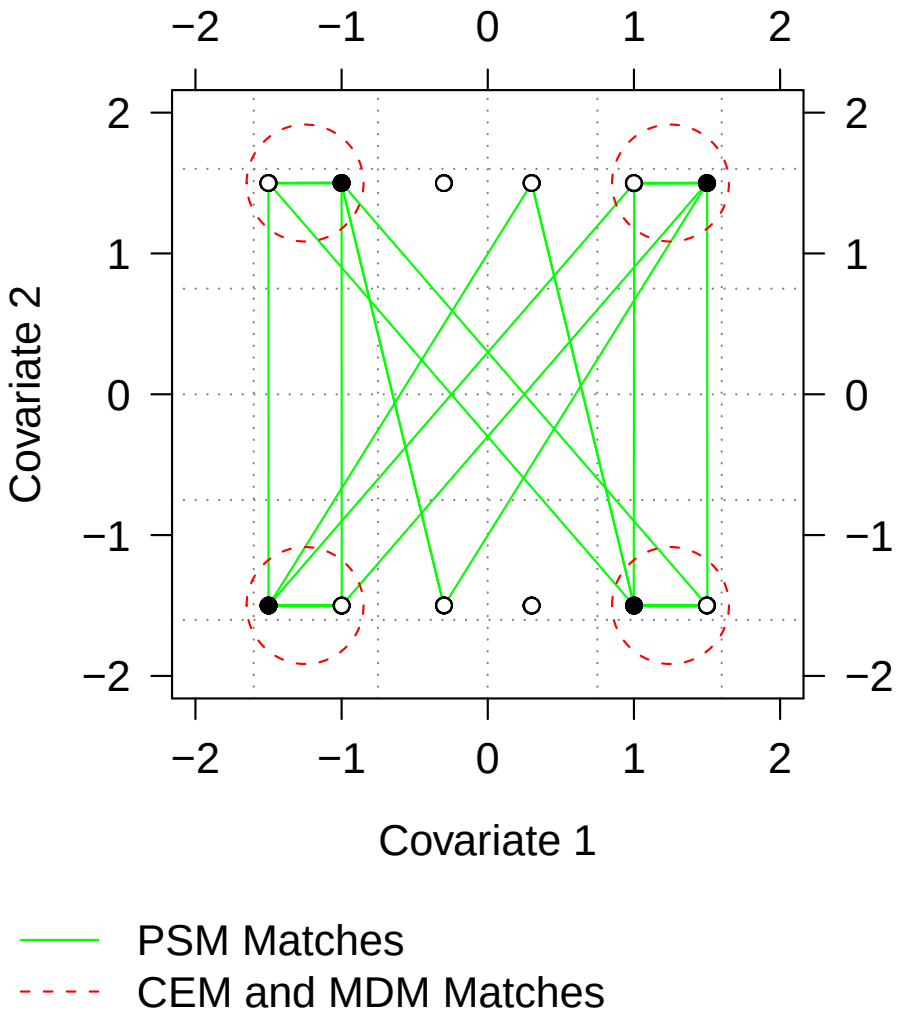
Problems with PSM: Healthways Data

King, Nielsen, Coberley, Pope, and Wells (2012)

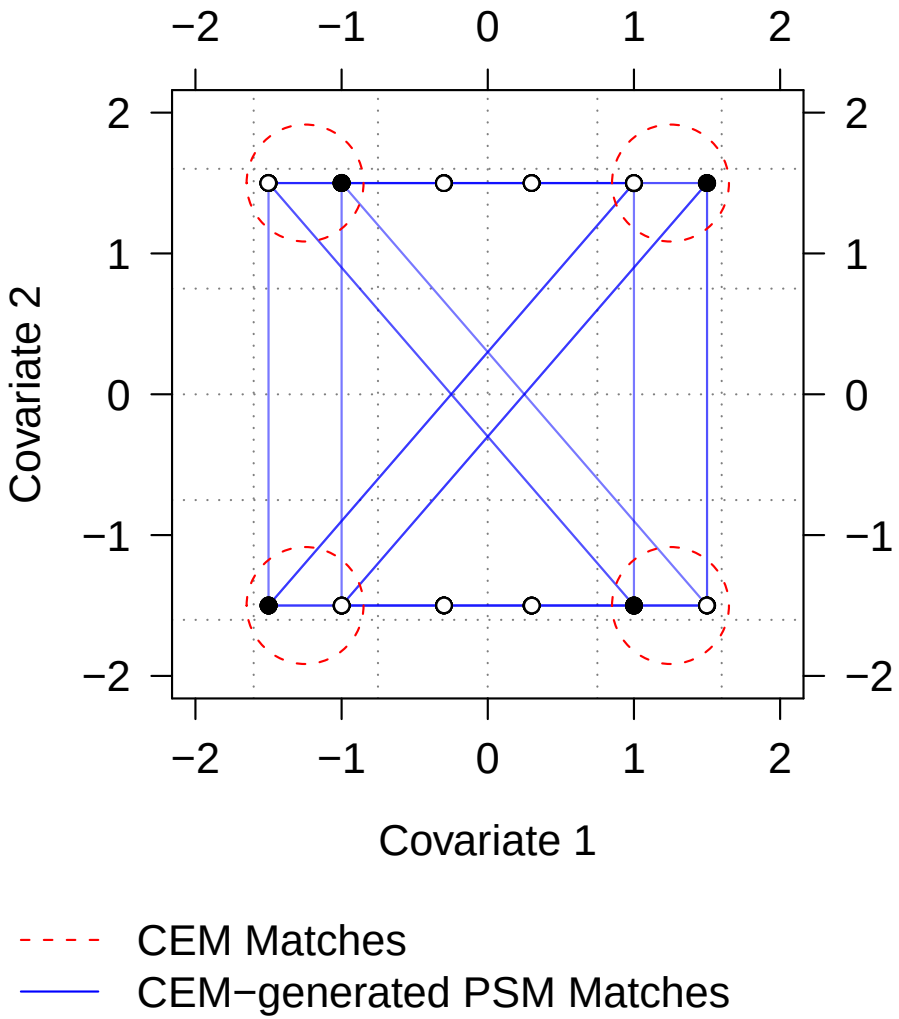


Methods-specific frontiers (for methodological research only)

PSM Approximates Random Matching in Balanced Data



Destroying CEM with PSM's Two Step Approach



Formal Notation and Assumptions for Causal Inference

Formal Notation and Assumptions for Causal Inference

- Units: $i = 1, \ldots, n$

Formal Notation and Assumptions for Causal Inference

- Units: $i = 1, \ldots, n$
- Treatment variable: $T_i = 1$ for treateds; 0 for controls

Formal Notation and Assumptions for Causal Inference

- Units: $i = 1, \ldots, n$
- Treatment variable: $T_i = 1$ for treateds; 0 for controls
- Potential outcomes: $Y_i(t)$ (for $t = 0, 1$) is the (potential) value of outcome variable if $T_i = t$

Formal Notation and Assumptions for Causal Inference

- Units: $i = 1, \ldots, n$
- Treatment variable: $T_i = 1$ for treateds; 0 for controls
- Potential outcomes: $Y_i(t)$ (for $t = 0, 1$) is the (potential) value of outcome variable if $T_i = t$
- Treatment effect: $\text{TE}_i = Y_i(1) - Y_i(0)$

Formal Notation and Assumptions for Causal Inference

- Units: $i = 1, \ldots, n$
- Treatment variable: $T_i = 1$ for treateds; 0 for controls
- Potential outcomes: $Y_i(t)$ (for $t = 0, 1$) is the (potential) value of outcome variable if $T_i = t$
- Treatment effect: $\text{TE}_i = Y_i(1) - Y_i(0)$
- Fundamental problem of causal inference: $Y_i(0)$ and $Y_i(1)$ are never both observed for any i

Formal Notation and Assumptions for Causal Inference

- Units: $i = 1, \ldots, n$
- Treatment variable: $T_i = 1$ for treateds; 0 for controls
- Potential outcomes: $Y_i(t)$ (for $t = 0, 1$) is the (potential) value of outcome variable if $T_i = t$
- Treatment effect: $\text{TE}_i = Y_i(1) - Y_i(0)$
- Fundamental problem of causal inference: $Y_i(0)$ and $Y_i(1)$ are never both observed for any i
- Observed outcome: $Y_i = T_i Y_i(1) + (1 - T_i) Y_i(0)$

Formal Notation and Assumptions for Causal Inference

- Units: $i = 1, \ldots, n$
- Treatment variable: $T_i = 1$ for treateds; 0 for controls
- Potential outcomes: $Y_i(t)$ (for $t = 0, 1$) is the (potential) value of outcome variable if $T_i = t$
- Treatment effect: $\text{TE}_i = Y_i(1) - Y_i(0)$
- Fundamental problem of causal inference: $Y_i(0)$ and $Y_i(1)$ are never both observed for any i
- Observed outcome: $Y_i = T_i Y_i(1) + (1 - T_i) Y_i(0)$
- Pretreatment control variables: X_i (k-vector)

Formal Notation and Assumptions for Causal Inference

- Units: $i = 1, \ldots, n$
- Treatment variable: $T_i = 1$ for treateds; 0 for controls
- Potential outcomes: $Y_i(t)$ (for $t = 0, 1$) is the (potential) value of outcome variable if $T_i = t$
- Treatment effect: $\text{TE}_i = Y_i(1) - Y_i(0)$
- Fundamental problem of causal inference: $Y_i(0)$ and $Y_i(1)$ are never both observed for any i
- Observed outcome: $Y_i = T_i Y_i(1) + (1 - T_i) Y_i(0)$
- Pretreatment control variables: X_i (k-vector)
- Simplification we will use: focus on TE for treated units; so $Y_i(0)$ is unobserved; $Y_i(1) = Y_i$

Implied Notation Assumptions

- Overlap (aka common support): $\Pr(T_i = 1|X) < 1$ for all treated i

Implied Notation Assumptions

- Overlap (aka common support): $\Pr(T_i = 1|X) < 1$ for all treated i
- If no overlap: TE_i does not logically exist

Implied Notation Assumptions

- Overlap (aka common support): $\Pr(T_i = 1|X) < 1$ for all treated i
- If no overlap: TE_i does not logically exist
- Stable Unit Treatment Value (SUTVA): logical consistency, so potential values are fixed if T changes (or "no interference"; no "versions of treatment").

Implied Notation Assumptions

- Overlap (aka common support): $\Pr(T_i = 1|X) < 1$ for all treated i
- If no overlap: TE_i does not logically exist
- Stable Unit Treatment Value (SUTVA): logical consistency, so potential values are fixed if T changes (or "no interference"; no "versions of treatment").
- Example of no SUTA: $Y_i(0) = 5$ if $T_i = 1$ but $Y_i(0) = 8$ if it were the case that $T_i = 0$

Implied Notation Assumptions

- Overlap (aka common support): $\Pr(T_i = 1|X) < 1$ for all treated i
- If no overlap: TE_i does not logically exist
- Stable Unit Treatment Value (SUTVA): logical consistency, so potential values are fixed if T changes (or "no interference"; no "versions of treatment").
- Example of no SUTA: $Y_i(0) = 5$ if $T_i = 1$ but $Y_i(0) = 8$ if it were the case that $T_i = 0$
- If overlap, but no SUTVA: $Y_i(0)$ and TE_i exist but are not fixed QOIs

Causal Quantities of Interest

Causal Quantities of Interest

- Sample Average Treatment Effect on the Treated

$$\text{SATT} = \text{mean}_{i \in \{T=1\}}(\text{TE}_i)$$

Causal Quantities of Interest

- Sample Average Treatment Effect on the Treated

$$\text{SATT} = \text{mean}_{i \in \{T=1\}}(\text{TE}_i)$$

- Feasible SATT

FSATT: TE_i averaged over well-matched treated units

Causal Quantities of Interest

- Sample Average Treatment Effect on the Treated

$$\text{SATT} = \text{mean}_{i \in \{T=1\}}(\text{TE}_i)$$

- Feasible SATT

FSATT: TE_i averaged over well-matched treated units

- Other QOIs: PATT, FPATT; SATE, FSATE; PATE, FPATE

Causal Estimation Assumptions

Causal Estimation Assumptions

- Ignorable Treatment Assignment (ITA)

Causal Estimation Assumptions

- Ignorable Treatment Assignment (ITA)
  - Goal: Identification (we can learn the truth if $n \to \infty$)

Causal Estimation Assumptions

- Ignorable Treatment Assignment (ITA)
  - Goal: Identification (we can learn the truth if $n \to \infty$)
  - The mechanism that produces treatment assignment (T_i) is independent of the potential outcomes

Causal Estimation Assumptions

- Ignorable Treatment Assignment (ITA)
  - Goal: Identification (we can learn the truth if $n \to \infty$)
  - The mechanism that produces treatment assignment (T_i) is independent of the potential outcomes
  - Formally: $T_i \perp Y_i(0)|X$ for all treateds

Causal Estimation Assumptions

- Ignorable Treatment Assignment (ITA)
  - Goal: Identification (we can learn the truth if $n \to \infty$)
  - The mechanism that produces treatment assignment (T_i) is independent of the potential outcomes
  - Formally: $T_i \perp Y_i(0) | X$ for all treateds
  - aka: "selection on observables," "no selection on Y," "unconfoundedness," "conditional independence"; special cases: "exogeneity," "no omitted variable bias"

Causal Estimation Assumptions

- Ignorable Treatment Assignment (ITA)
  - Goal: Identification (we can learn the truth if $n \to \infty$)
  - The mechanism that produces treatment assignment (T_i) is independent of the potential outcomes
  - Formally: $T_i \perp Y_i(0)|X$ for all treateds
  - aka: "selection on observables," "no selection on Y," "unconfoundedness," "conditional independence"; special cases: "exogeneity," "no omitted variable bias"
- Uncorrelated Treatment Assignment (UTA) (ITA Simplified)

Causal Estimation Assumptions

- Ignorable Treatment Assignment (ITA)
  - Goal: Identification (we can learn the truth if $n \to \infty$)
  - The mechanism that produces treatment assignment (T_i) is independent of the potential outcomes
  - Formally: $T_i \perp Y_i(0)|X$ for all treateds
  - aka: "selection on observables," "no selection on Y," "unconfoundedness," "conditional independence"; special cases: "exogeneity," "no omitted variable bias"
- Uncorrelated Treatment Assignment (UTA) (ITA Simplified)
  - Goal: $\widehat{\text{SATT}} = \text{SATT}$

Causal Estimation Assumptions

- Ignorable Treatment Assignment (ITA)
  - Goal: Identification (we can learn the truth if $n \to \infty$)
  - The mechanism that produces treatment assignment (T_i) is independent of the potential outcomes
  - Formally: $T_i \perp Y_i(0)|X$ for all treateds
  - aka: "selection on observables," "no selection on Y," "unconfoundedness," "conditional independence"; special cases: "exogeneity," "no omitted variable bias"
- Uncorrelated Treatment Assignment (UTA) (ITA Simplified)
  - Goal: $\widehat{\text{SATT}} = \text{SATT}$
  - Sufficient (not necess.): $(Y_i(0)|T_i = 1, X) = (Y_j(0)|T_j = 0, X)$ $\forall$ treateds i & matching controls j

Causal Estimation Assumptions

- Ignorable Treatment Assignment (ITA)
  - Goal: Identification (we can learn the truth if $n \to \infty$)
  - The mechanism that produces treatment assignment (T_i) is independent of the potential outcomes
  - Formally: $T_i \perp Y_i(0)|X$ for all treateds
  - aka: "selection on observables," "no selection on Y," "unconfoundedness," "conditional independence"; special cases: "exogeneity," "no omitted variable bias"
- Uncorrelated Treatment Assignment (UTA) (ITA Simplified)
  - Goal: $\widehat{\text{SATT}} = \text{SATT}$
  - Sufficient (not necess.): $(Y_i(0)|T_i = 1, X) = (Y_j(0)|T_j = 0, X)$ $\forall$ treateds i & matching controls j
  - Necessary: $\text{mean}(Y_i(0)|T = 0, X) = \text{mean}(Y_j(0)|T = 1, X)$ (average within strata of X are equal)

Causal Estimation Assumptions

- Ignorable Treatment Assignment (ITA)
  - Goal: Identification (we can learn the truth if $n \to \infty$)
  - The mechanism that produces treatment assignment (T_i) is independent of the potential outcomes
  - Formally: $T_i \perp Y_i(0)|X$ for all treateds
  - aka: "selection on observables," "no selection on Y," "unconfoundedness," "conditional independence"; special cases: "exogeneity," "no omitted variable bias"
- Uncorrelated Treatment Assignment (UTA) (ITA Simplified)
  - Goal: $\widehat{\text{SATT}} = \text{SATT}$
  - Sufficient (not necess.): $(Y_i(0)|T_i = 1, X) = (Y_j(0)|T_j = 0, X)$ $\forall$ treateds i & matching controls j
  - Necessary: $\text{mean}(Y_i(0)|T = 0, X) = \text{mean}(Y_j(0)|T = 1, X)$ (average within strata of X are equal)
  - Simpler special case under 1-to-1 matching on X: $\text{mean}(Y_i(0)|T_i = 1) = \text{mean}(Y_j|T_j = 0)$

Conclusions

Conclusions

- The Matching Frontier

Conclusions

- The Matching Frontier
  - Fast; easy to use; no need to iterate

Conclusions

- The Matching Frontier
  - Fast; easy to use; no need to iterate
  - No need to choose among matching methods

Conclusions

- The Matching Frontier
  - Fast; easy to use; no need to iterate
  - No need to choose among matching methods
  - Optimal results for your choice of imbalance metric

Conclusions

- The Matching Frontier
  - Fast; easy to use; no need to iterate
  - No need to choose among matching methods
  - Optimal results for your choice of imbalance metric
- Propensity score matching:

Conclusions

- The Matching Frontier
  - Fast; easy to use; no need to iterate
  - No need to choose among matching methods
  - Optimal results for your choice of imbalance metric
- Propensity score matching:
  - The problem:

Conclusions

- The Matching Frontier
  - Fast; easy to use; no need to iterate
  - No need to choose among matching methods
  - Optimal results for your choice of imbalance metric
- Propensity score matching:
  - The problem:
    - Imbalance can be worse than original data

Conclusions

- The Matching Frontier
  - Fast; easy to use; no need to iterate
  - No need to choose among matching methods
  - Optimal results for your choice of imbalance metric
- Propensity score matching:
  - The problem:
    - Imbalance can be worse than original data
    - Can increase imbalance when removing the worst matches

Conclusions

- The Matching Frontier
  - Fast; easy to use; no need to iterate
  - No need to choose among matching methods
  - Optimal results for your choice of imbalance metric
- Propensity score matching:
  - The problem:
    - Imbalance can be worse than original data
    - Can increase imbalance when removing the worst matches
    - Approximates random matching in well-balanced data (Random matching increases imbalance)

Conclusions

- The Matching Frontier
  - Fast; easy to use; no need to iterate
  - No need to choose among matching methods
  - Optimal results for your choice of imbalance metric
- Propensity score matching:
  - The problem:
    - Imbalance can be worse than original data
    - Can increase imbalance when removing the worst matches
    - Approximates random matching in well-balanced data (Random matching increases imbalance)
  - Implications:

Conclusions

- The Matching Frontier
  - Fast; easy to use; no need to iterate
  - No need to choose among matching methods
  - Optimal results for your choice of imbalance metric
- Propensity score matching:
  - The problem:
    - Imbalance can be worse than original data
    - Can increase imbalance when removing the worst matches
    - Approximates random matching in well-balanced data (Random matching increases imbalance)
  - Implications:
    - Balance checking required

Conclusions

- The Matching Frontier
  - Fast; easy to use; no need to iterate
  - No need to choose among matching methods
  - Optimal results for your choice of imbalance metric
- Propensity score matching:
  - The problem:
    - Imbalance can be worse than original data
    - Can increase imbalance when removing the worst matches
    - Approximates random matching in well-balanced data (Random matching increases imbalance)
  - Implications:
    - Balance checking required
    - Adjusting for potentially irrelevant covariates *with PSM*: mistake

Conclusions

- The Matching Frontier
  - Fast; easy to use; no need to iterate
  - No need to choose among matching methods
  - Optimal results for your choice of imbalance metric
- Propensity score matching:
  - The problem:
    - Imbalance can be worse than original data
    - Can increase imbalance when removing the worst matches
    - Approximates random matching in well-balanced data (Random matching increases imbalance)
  - Implications:
    - Balance checking required
    - Adjusting for potentially irrelevant covariates *with PSM*: mistake
    - Adjusting experimental data *with PSM*: mistake

Conclusions

- The Matching Frontier
  - Fast; easy to use; no need to iterate
  - No need to choose among matching methods
  - Optimal results for your choice of imbalance metric
- Propensity score matching:
  - The problem:
    - Imbalance can be worse than original data
    - Can increase imbalance when removing the worst matches
    - Approximates random matching in well-balanced data (Random matching increases imbalance)
  - Implications:
    - Balance checking required
    - Adjusting for potentially irrelevant covariates *with PSM*: mistake
    - Adjusting experimental data *with PSM*: mistake
    - Reestimating the propensity score after eliminating noncommon support: mistake

Conclusions

- The Matching Frontier
  - Fast; easy to use; no need to iterate
  - No need to choose among matching methods
  - Optimal results for your choice of imbalance metric
- Propensity score matching:
  - The problem:
    - Imbalance can be worse than original data
    - Can increase imbalance when removing the worst matches
    - Approximates random matching in well-balanced data (Random matching increases imbalance)
  - Implications:
    - Balance checking required
    - Adjusting for potentially irrelevant covariates *with PSM*: mistake
    - Adjusting experimental data *with PSM*: mistake
    - Reestimating the propensity score after eliminating noncommon support: mistake
    - 1/4 caliper on propensity score: mistake

Conclusions

- The Matching Frontier
  - Fast; easy to use; no need to iterate
  - No need to choose among matching methods
  - Optimal results for your choice of imbalance metric
- Propensity score matching:
  - The problem:
    - Imbalance can be worse than original data
    - Can increase imbalance when removing the worst matches
    - Approximates random matching in well-balanced data (Random matching increases imbalance)
  - Implications:
    - Balance checking required
    - Adjusting for potentially irrelevant covariates *with PSM*: mistake
    - Adjusting experimental data *with PSM*: mistake
    - Reestimating the propensity score after eliminating noncommon support: mistake
    - 1/4 caliper on propensity score: mistake
- Software: MatchingFrontier