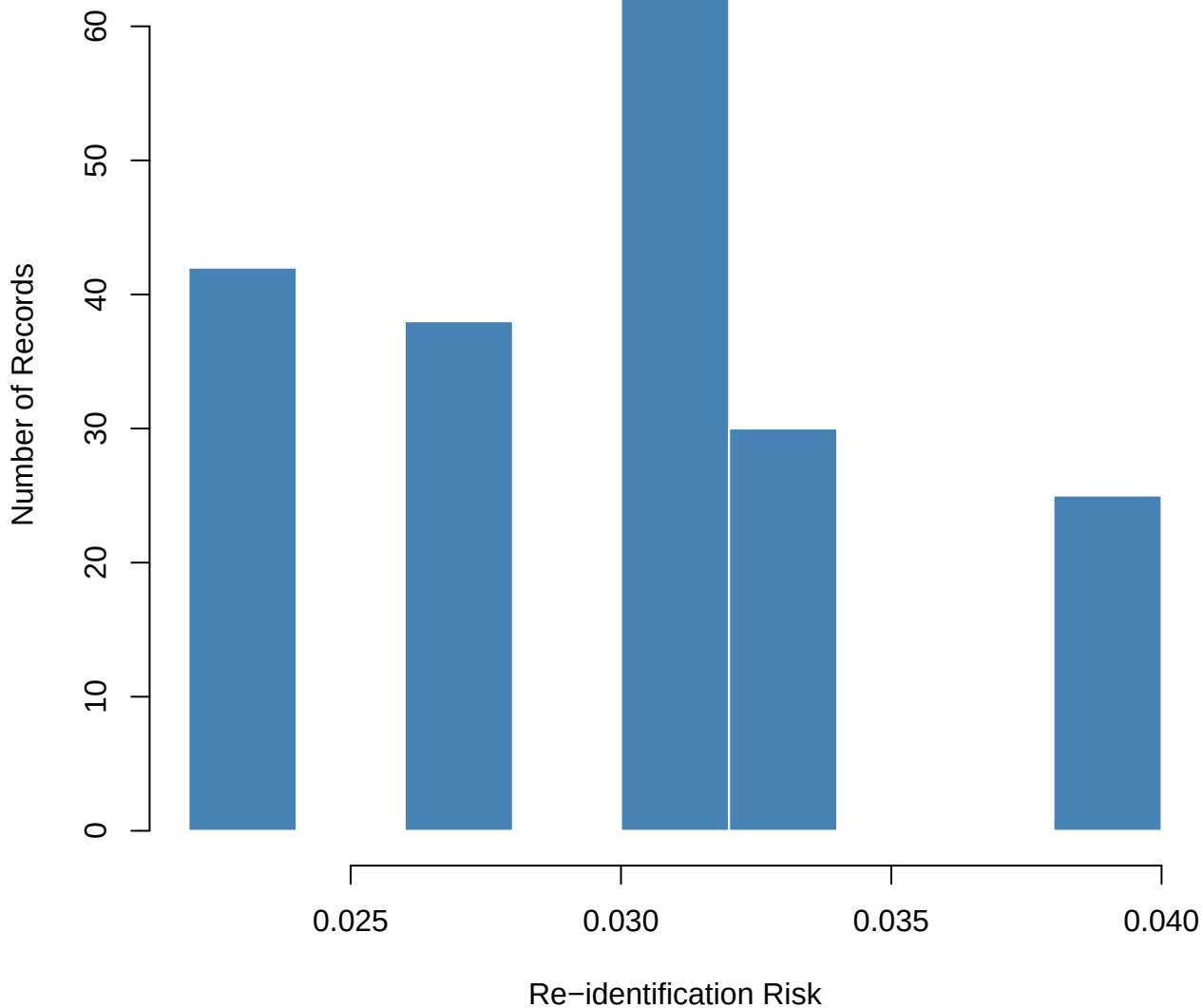
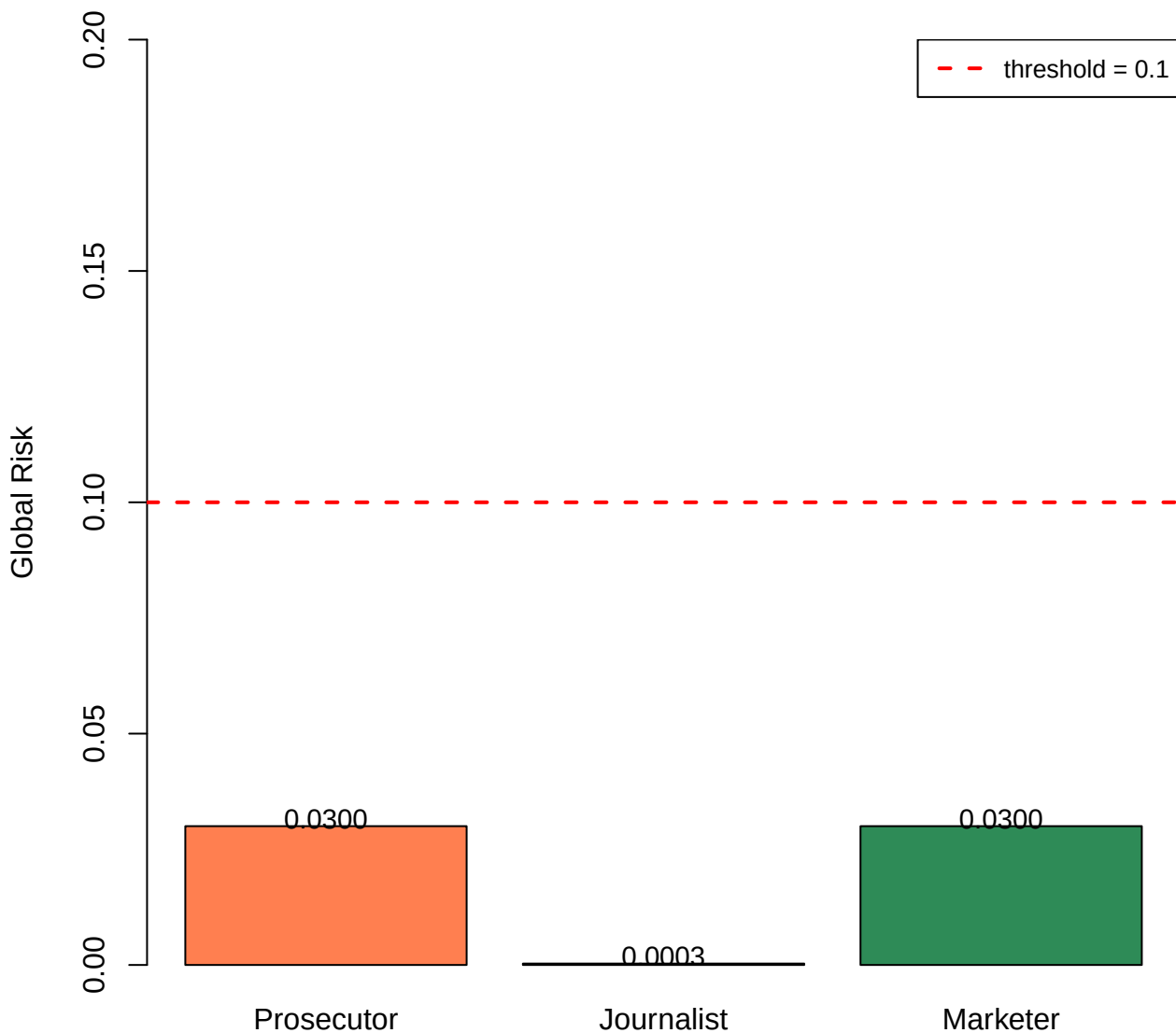


Per-Record Risk Distribution (Prosecutor Risk)

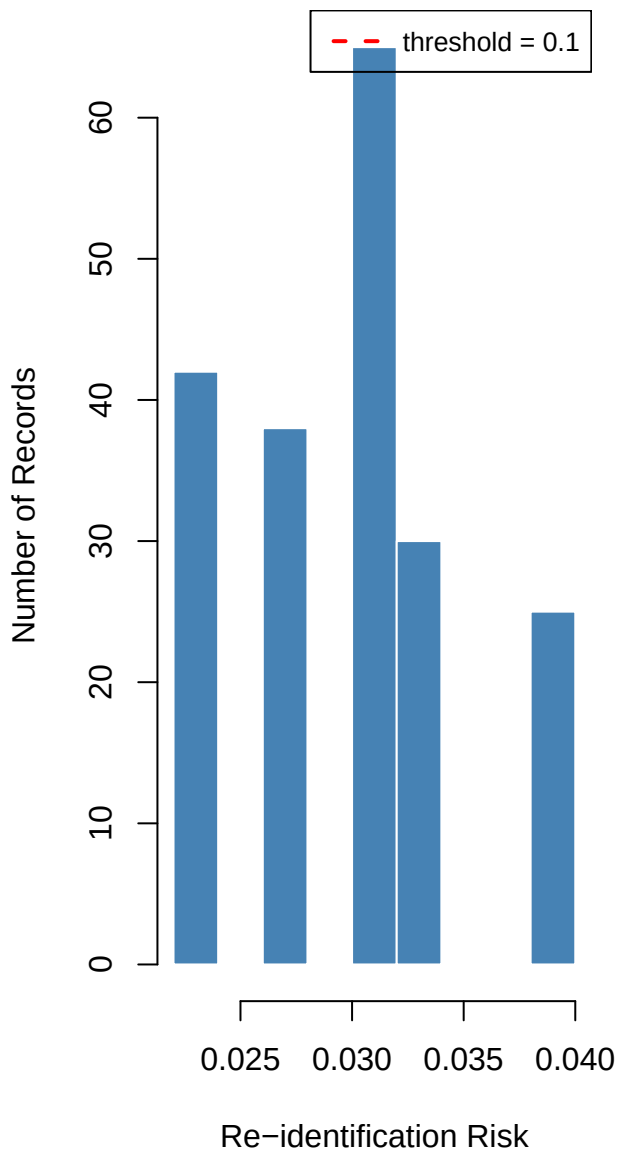
- - threshold = 0.1



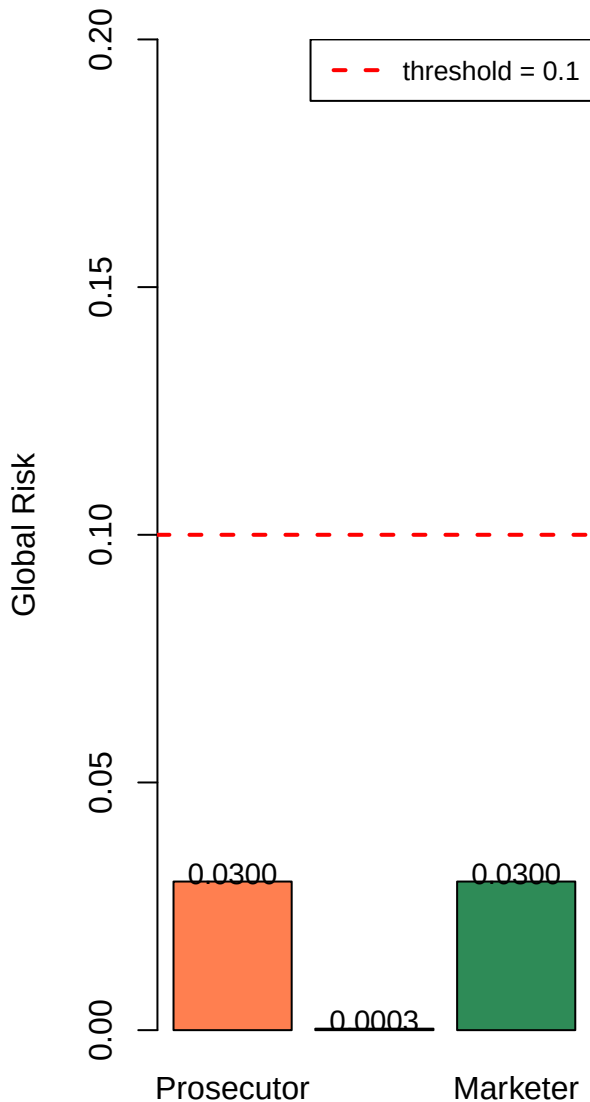
Global Risk by Attacker Model



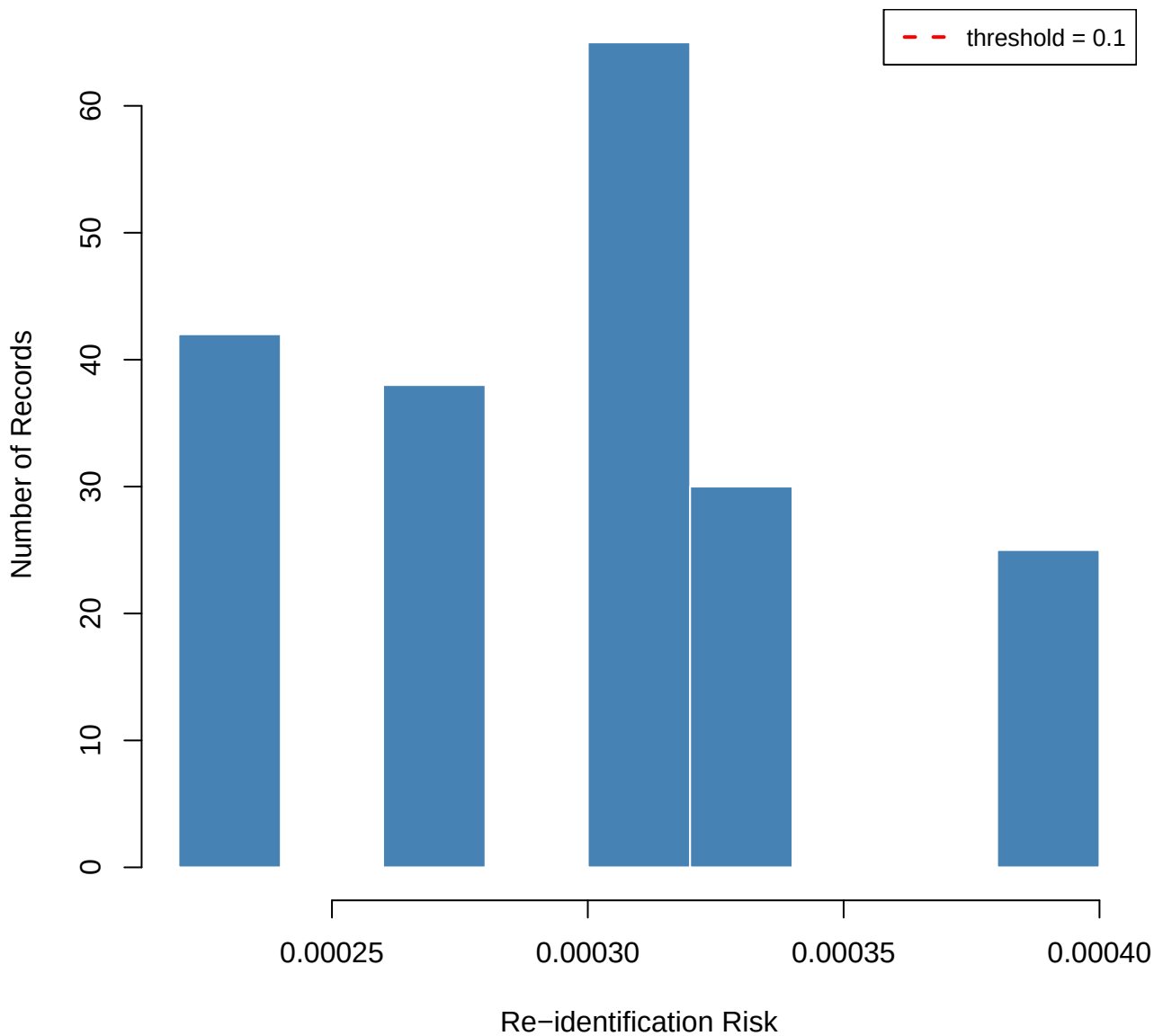
**Per-Record Risk Distribution
(Prosecutor Risk)**



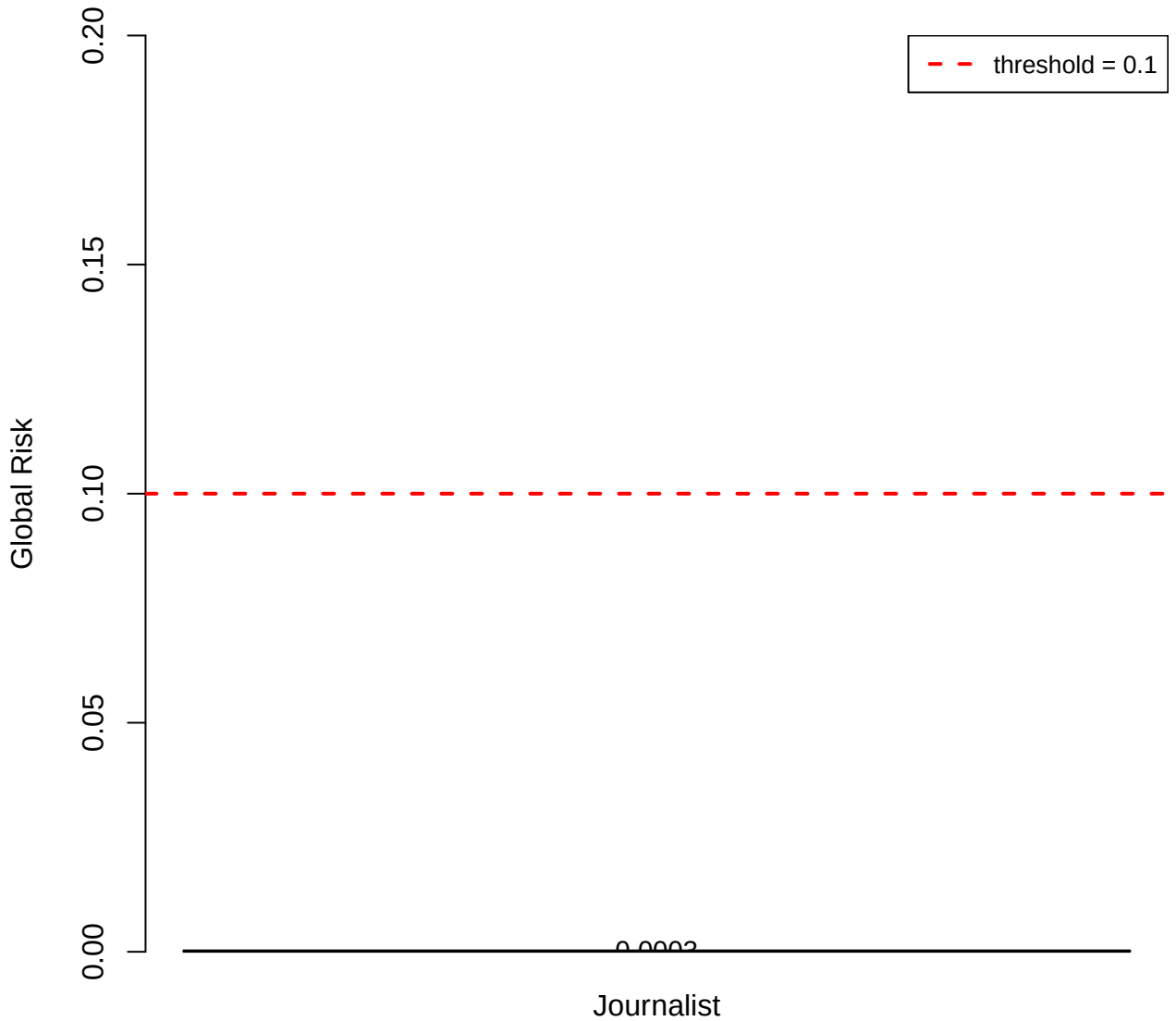
Global Risk by Attacker Model



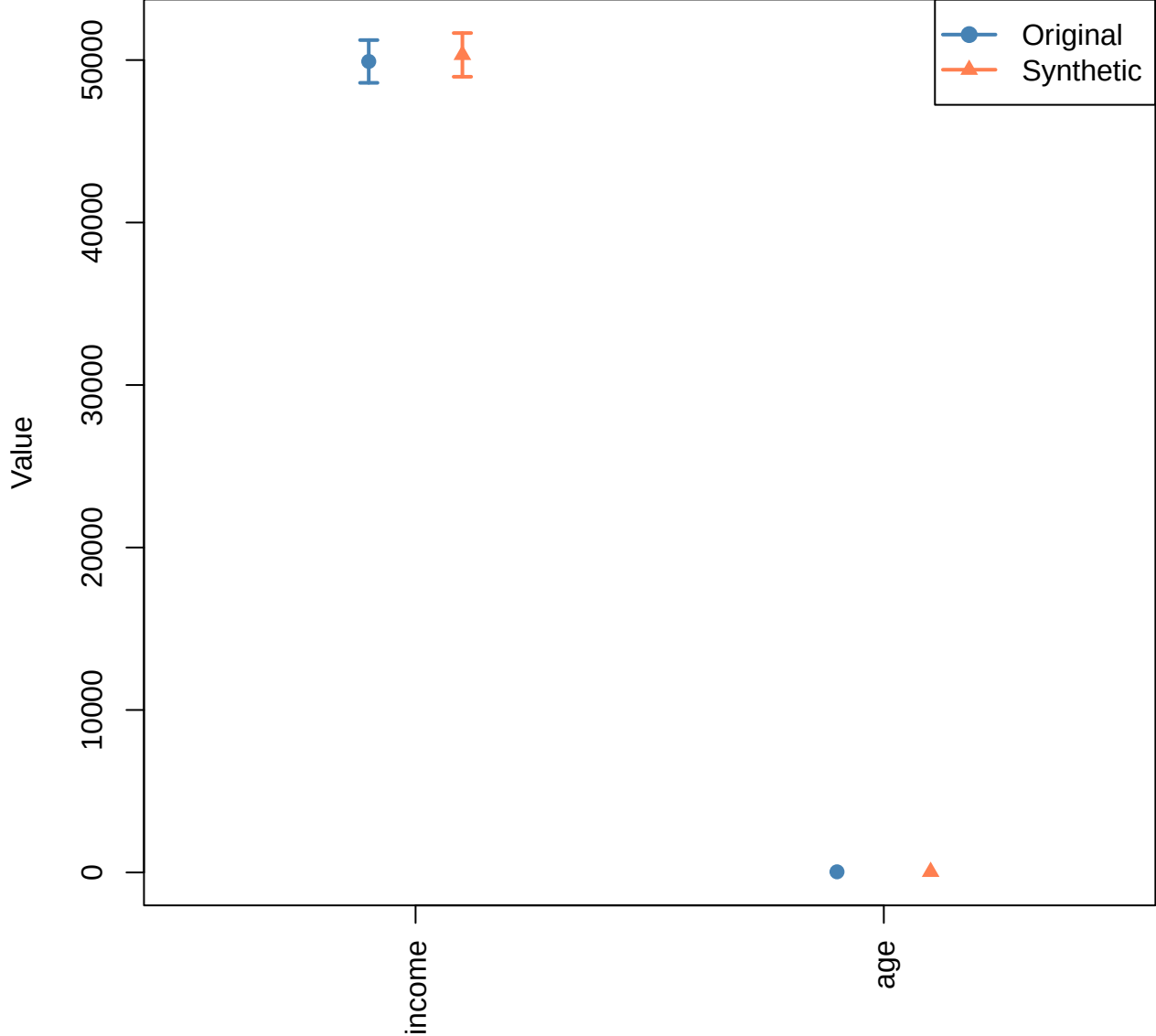
Per-Record Risk Distribution (Journalist Risk)



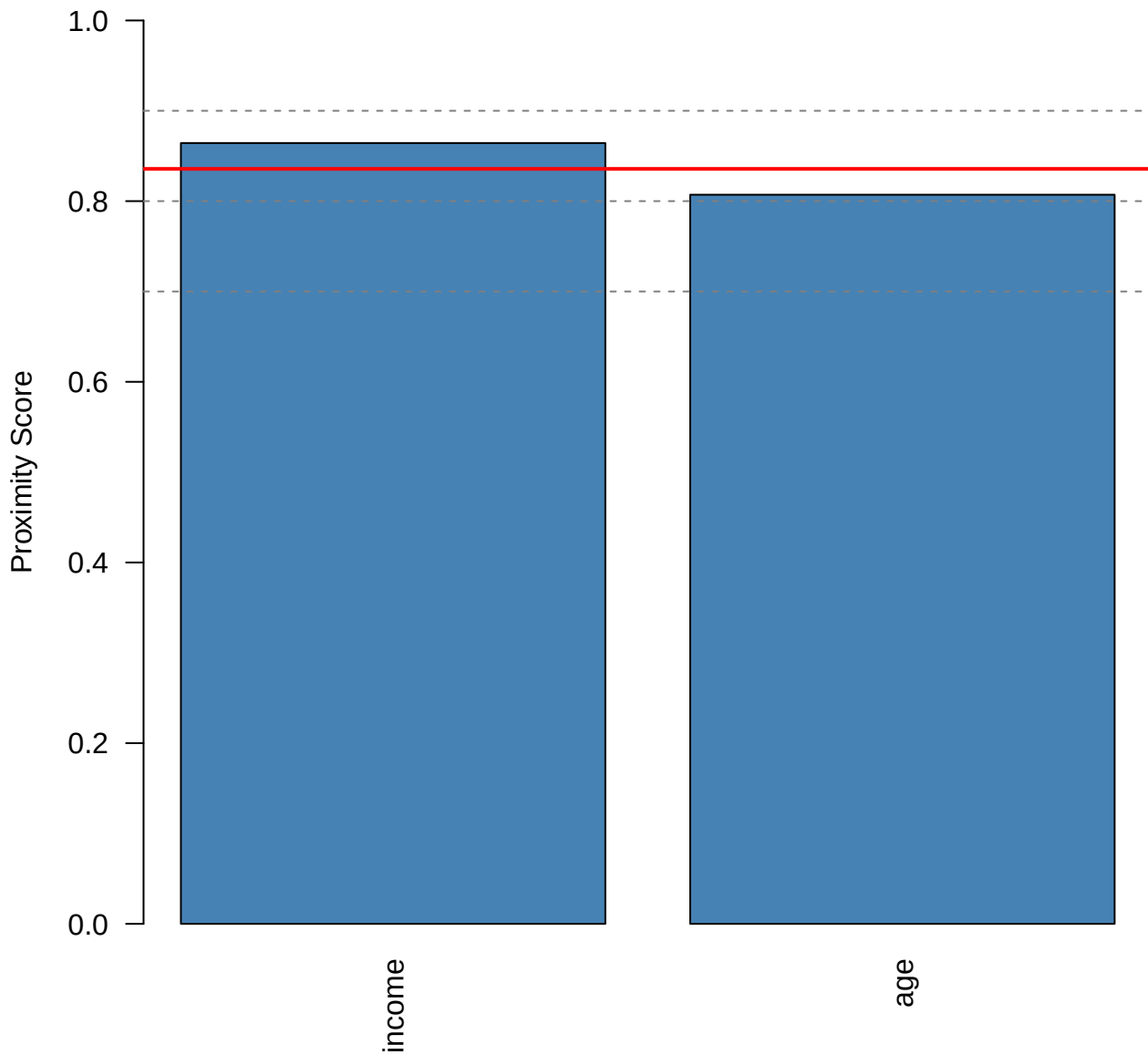
Global Risk by Attacker Model



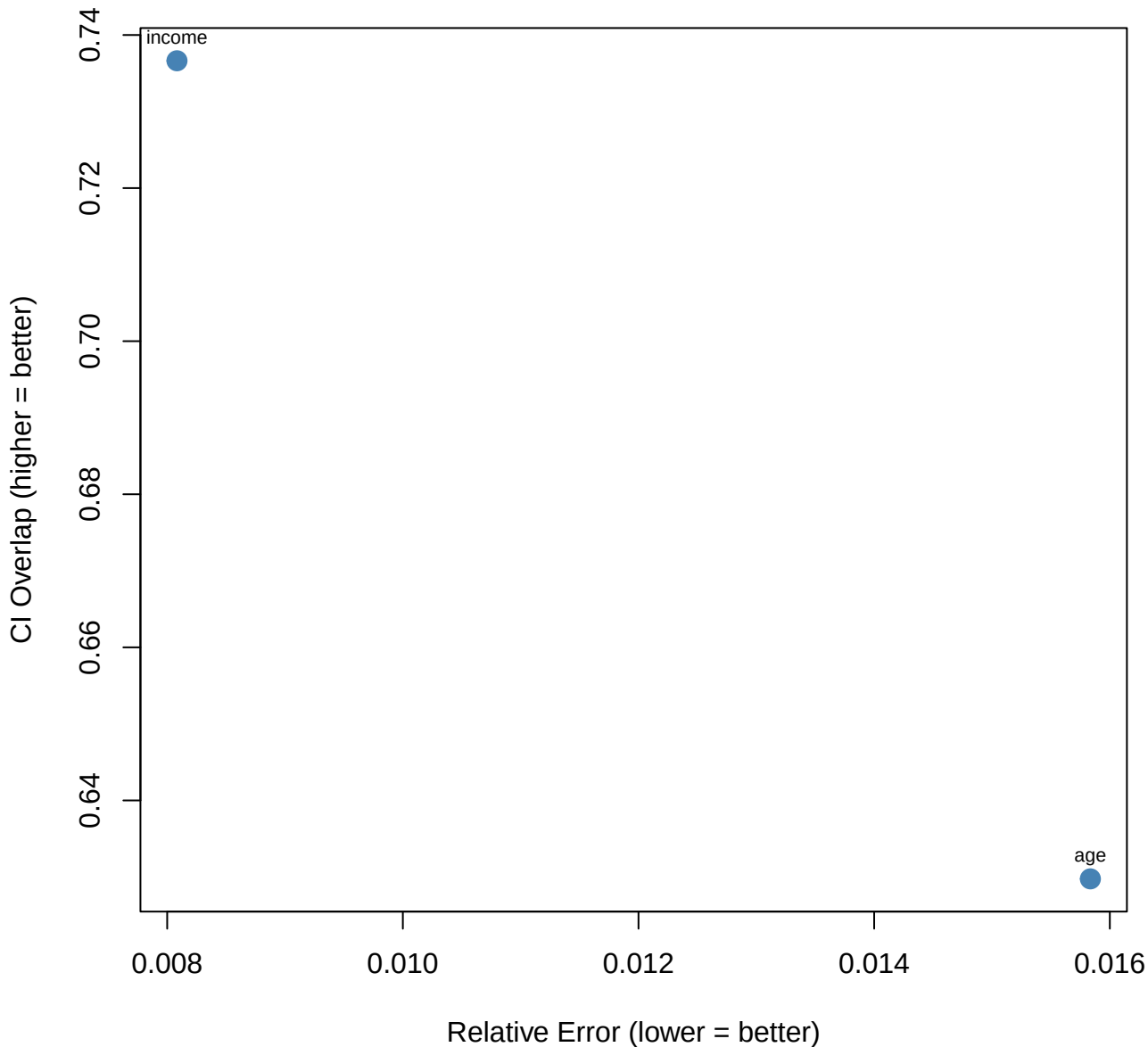
Confidence Interval Comparison



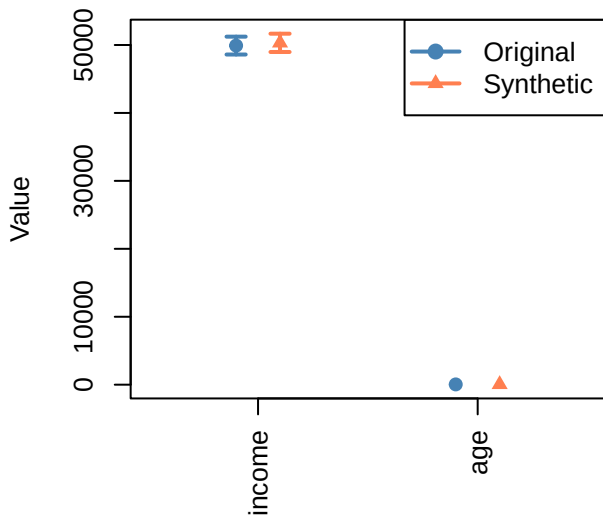
Proximity Scores by Variable



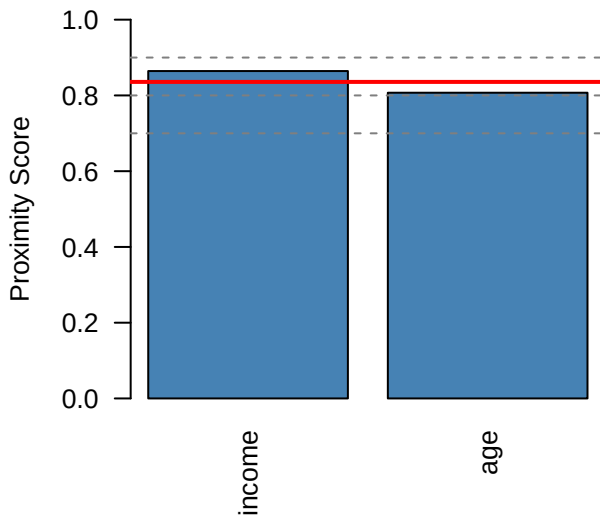
Error vs Overlap Trade-off



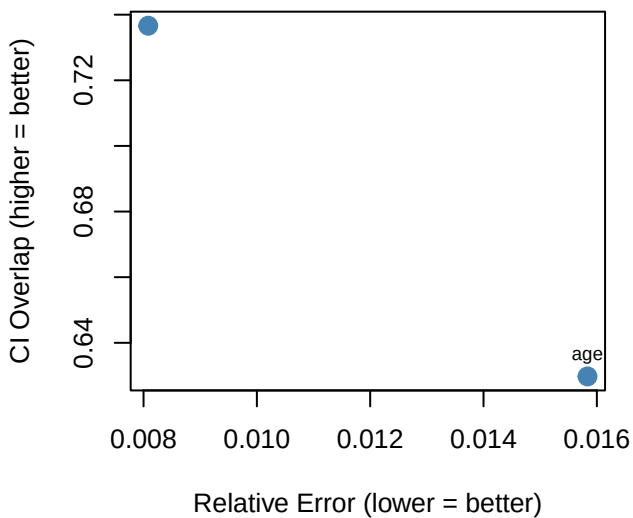
Confidence Interval Comparison



Proximity Scores by Variable

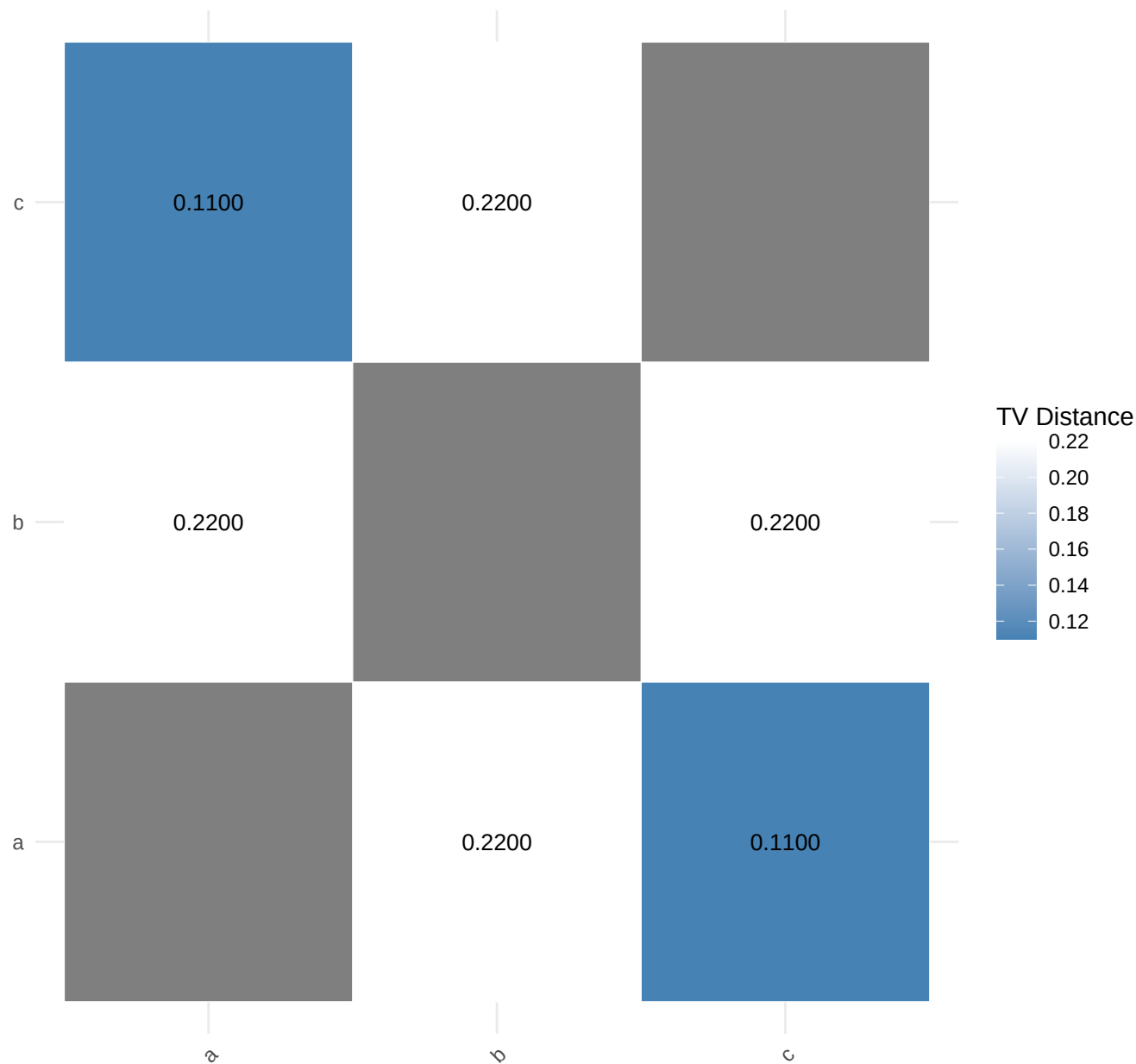


Error vs Overlap Trade-off



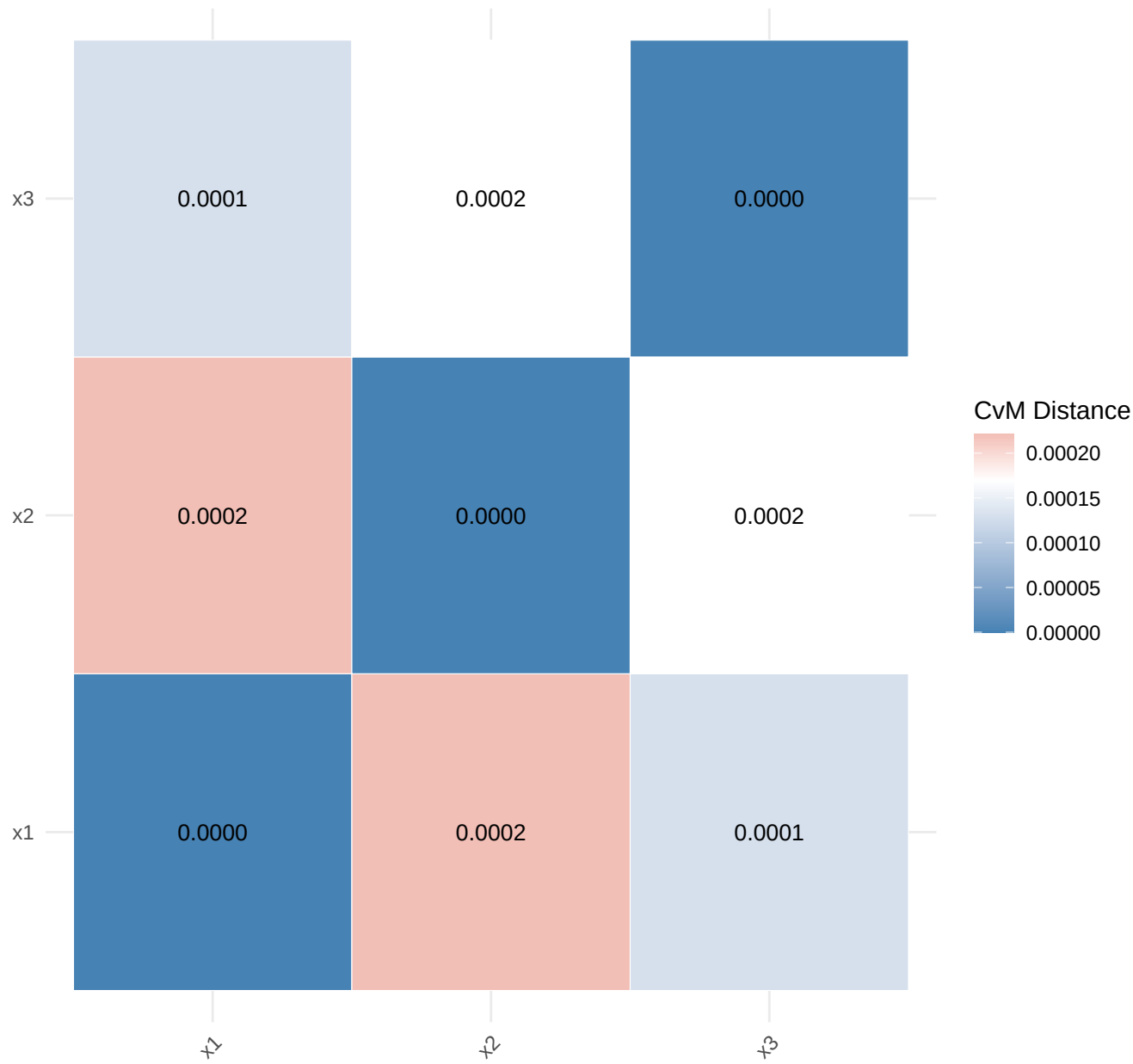
Pairwise Contingency Table TV Distances

Mean TV = 0.1833 | Utility = 0.8167

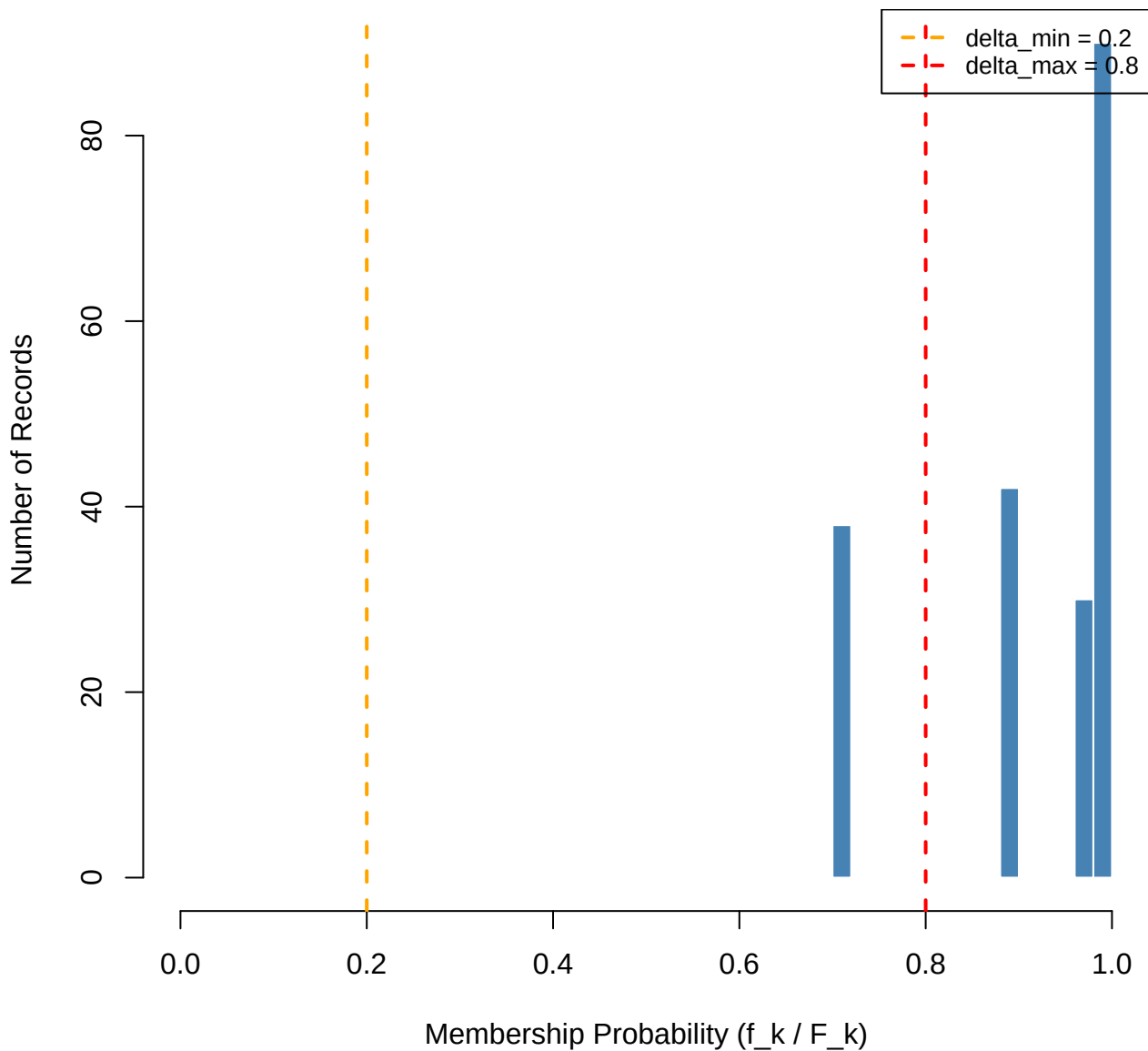


Pairwise Copula CvM Distances

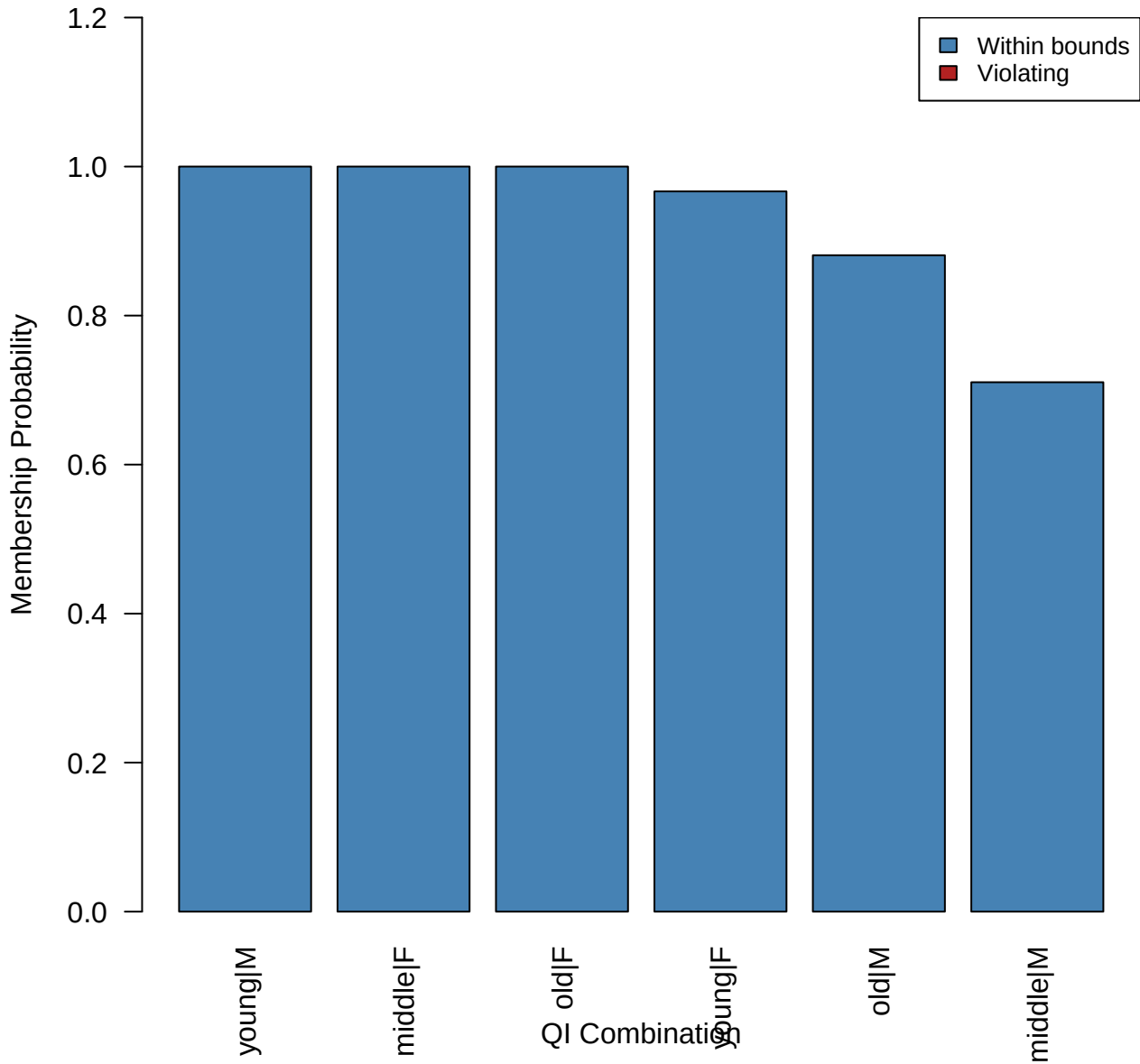
Mean CvM = 0.0002 | Utility = 0.9830



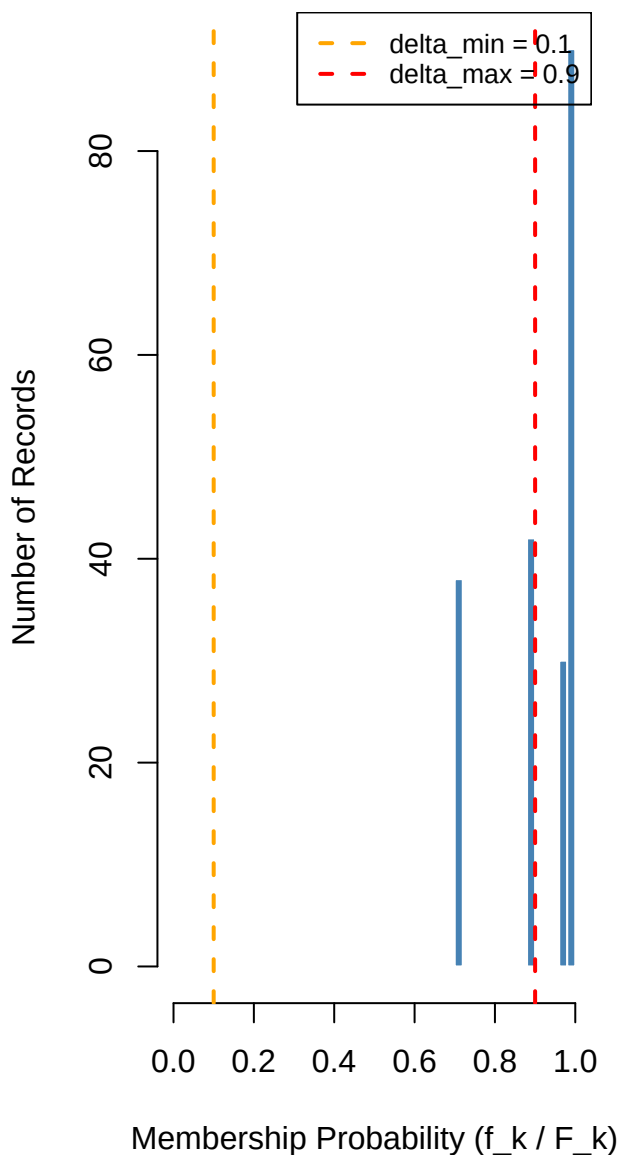
Membership Probability Distribution (0.2, 0.8)-presence



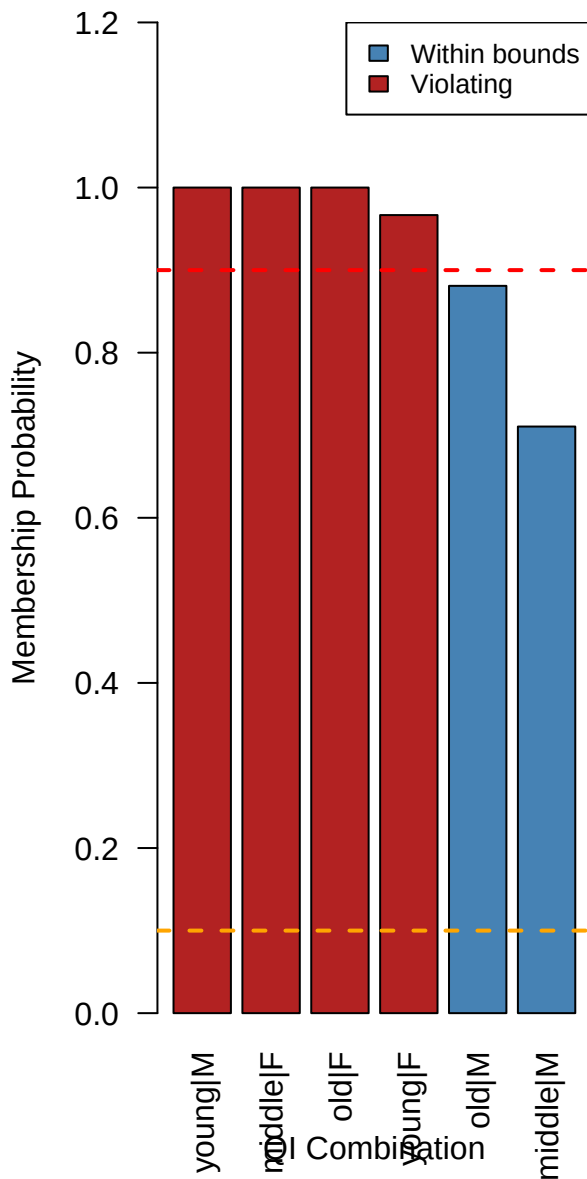
Membership Probability by QI Combination



**Membership Probability Distribution
(0.1, 0.9)-presence**

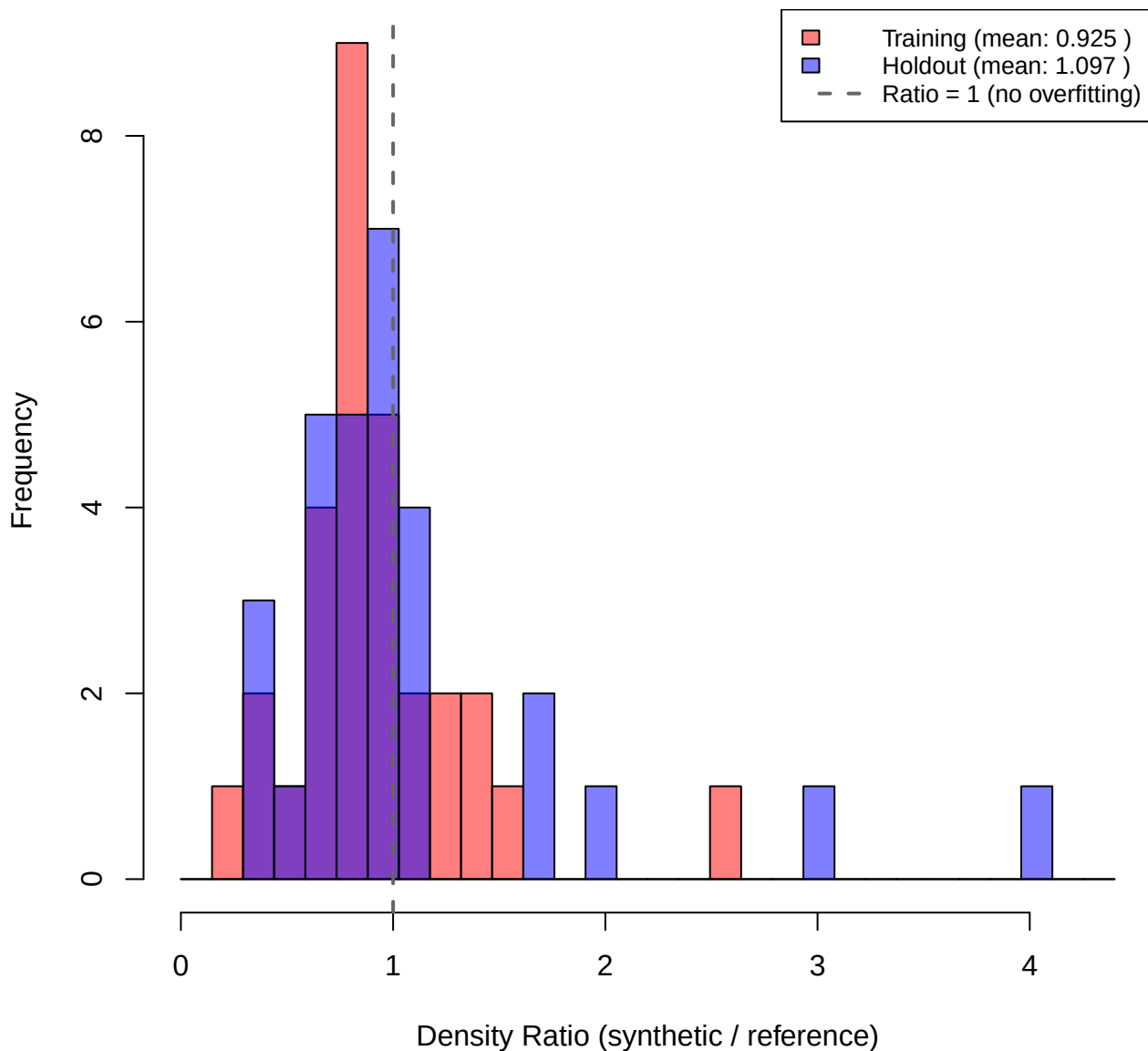


Membership Probability by QI Combination



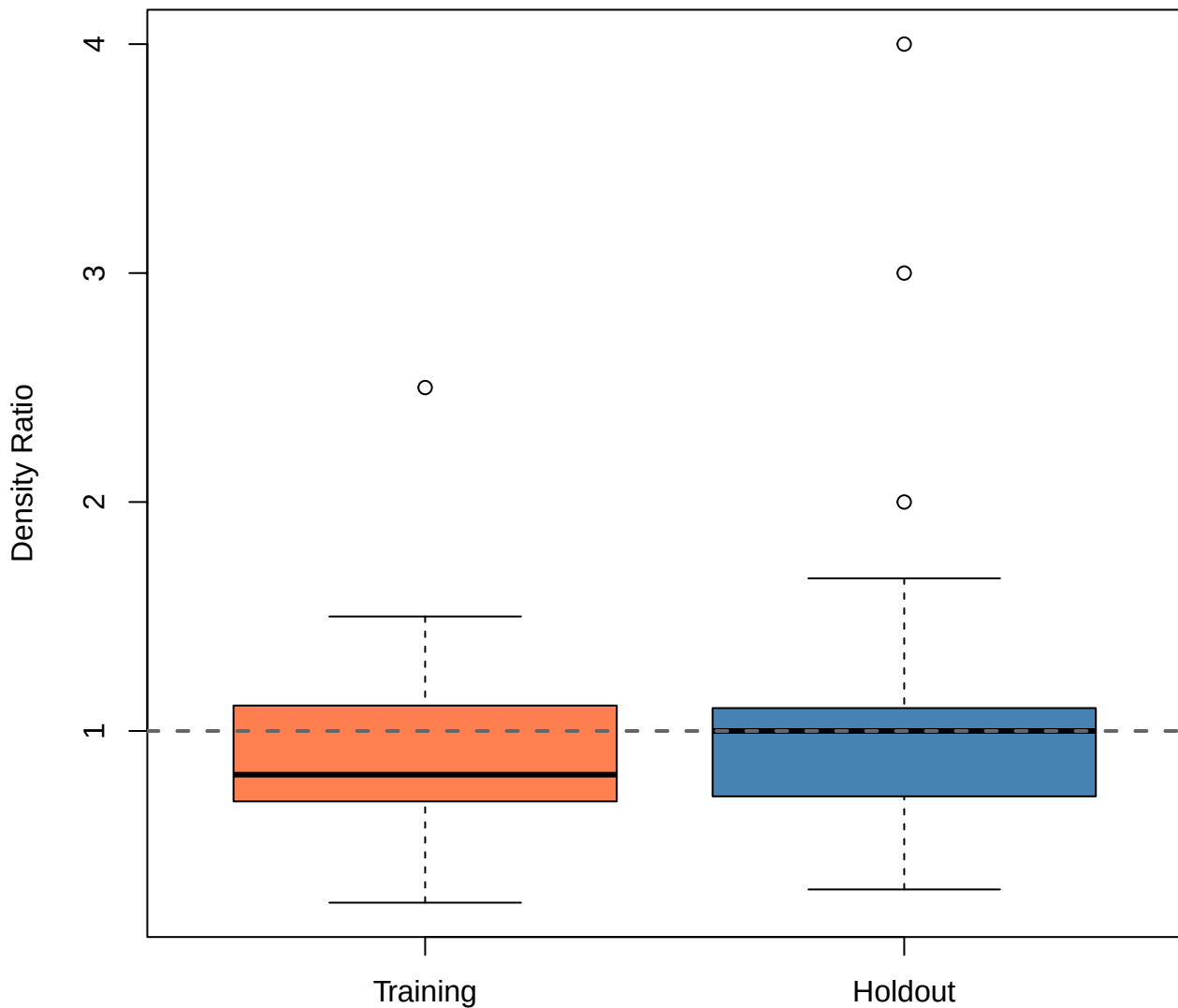
DOMIAS Density Ratios

AUC: 0.466



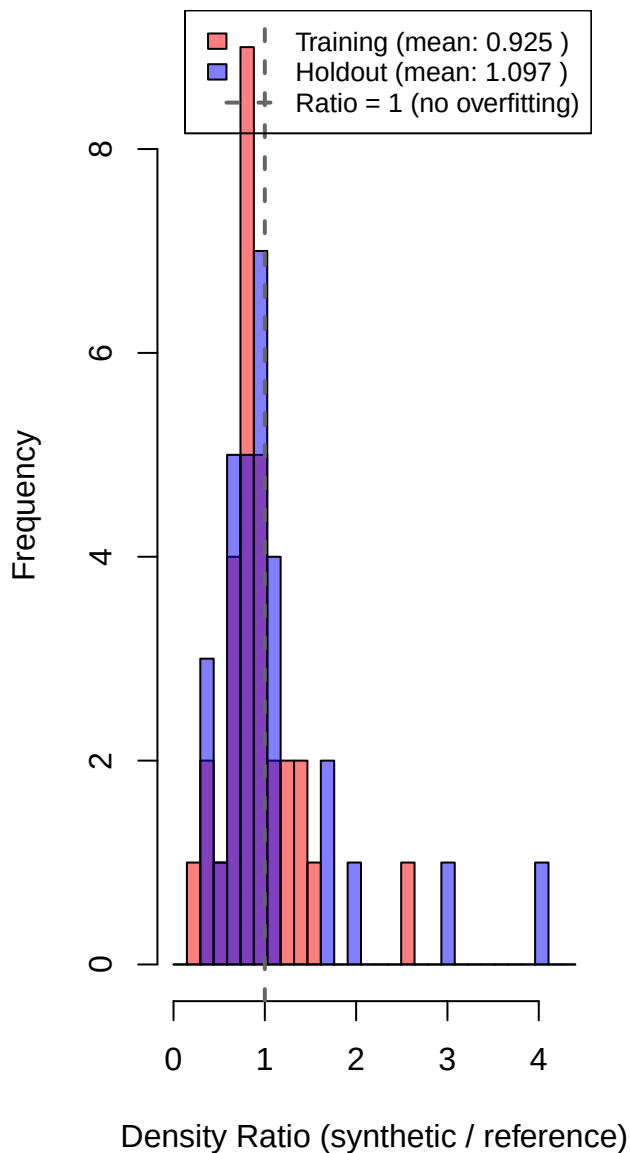
DOMIAS Density Ratios

Memorization Score: 0.843



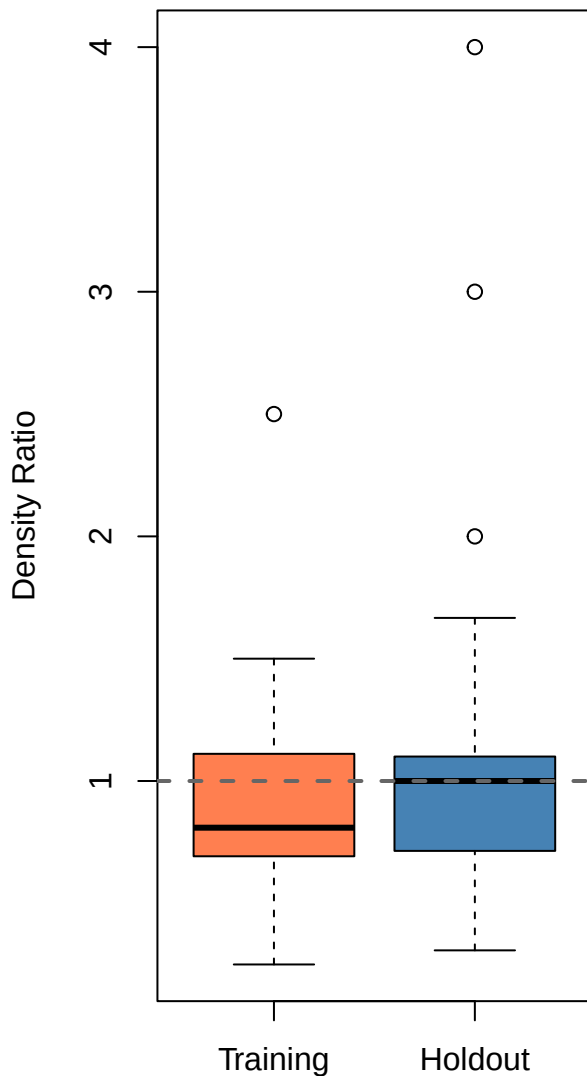
DOMIAS Density Ratios

AUC: 0.466

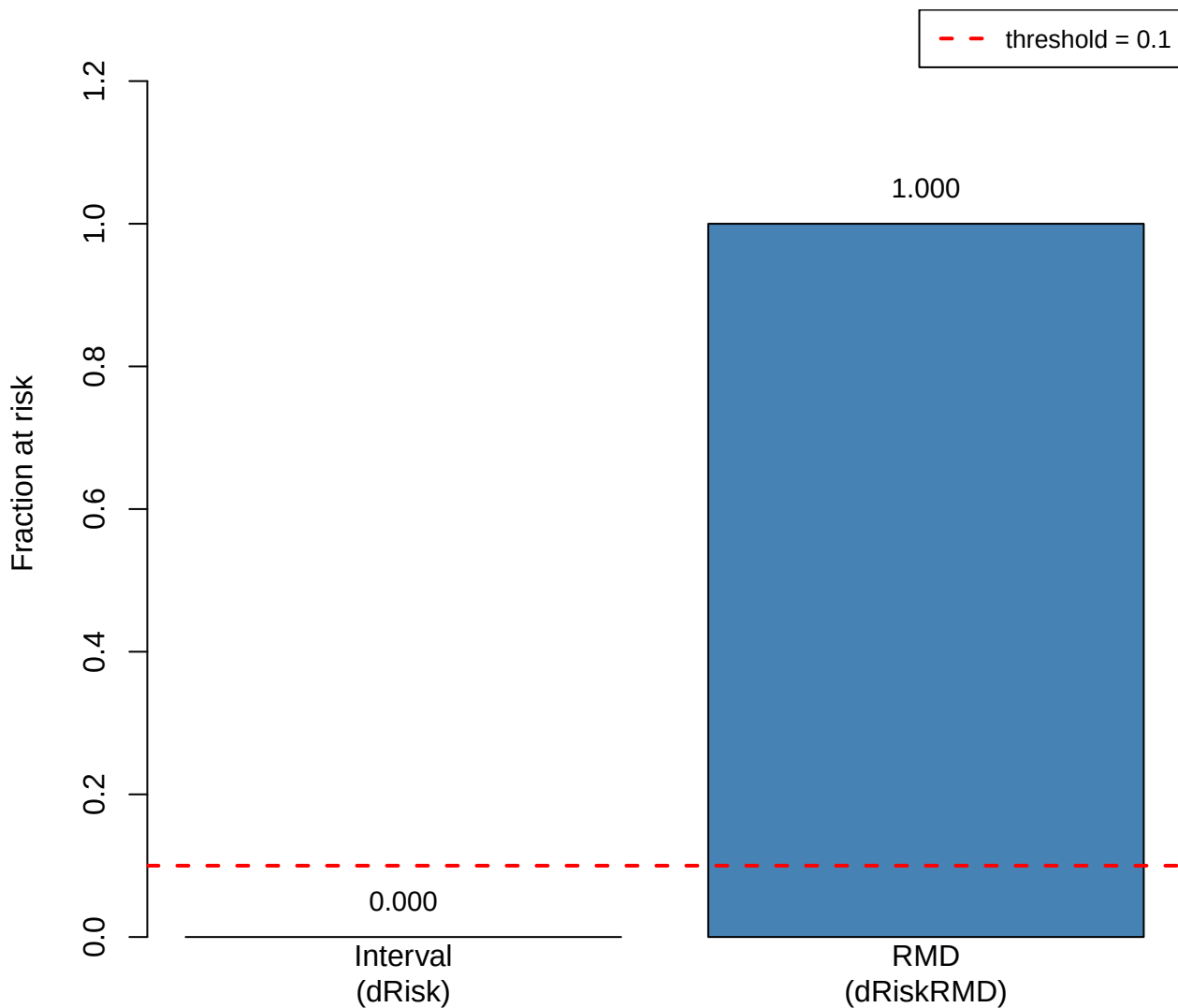


DOMIAS Density Ratios

Memorization Score: 0.843

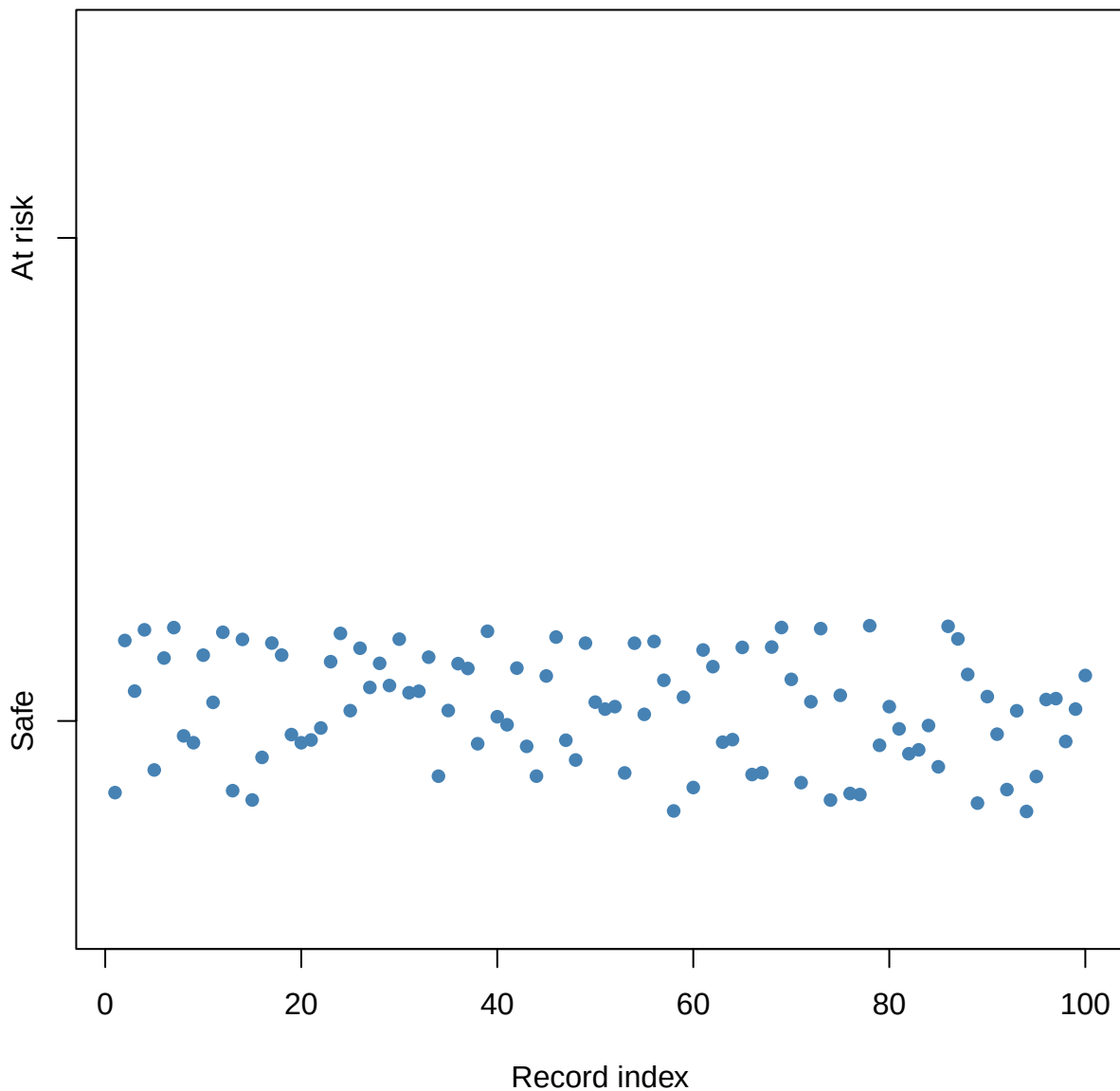


Disclosure Risk for Continuous Variables



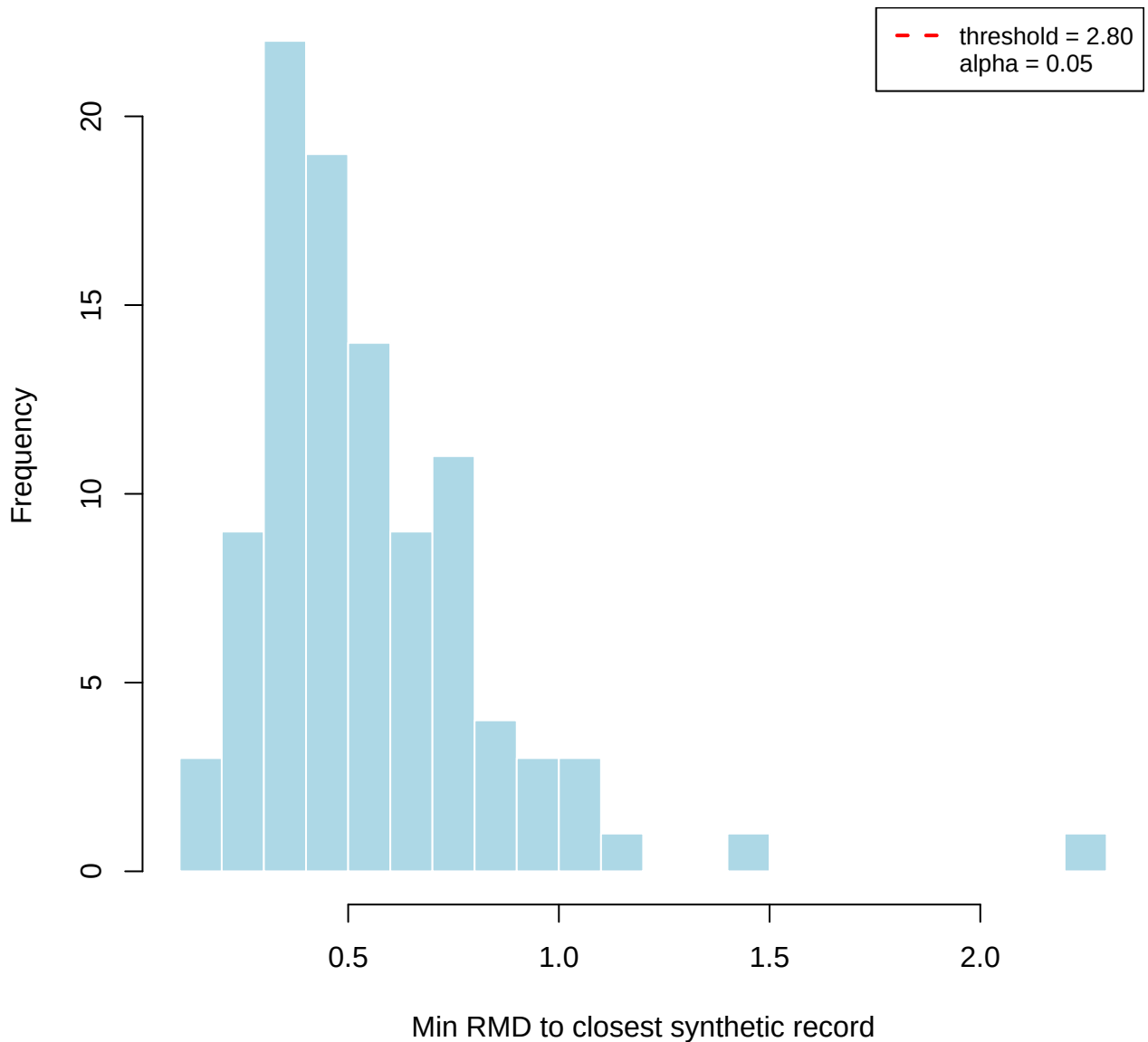
dRisk: Record Risk Status

0.0% at risk

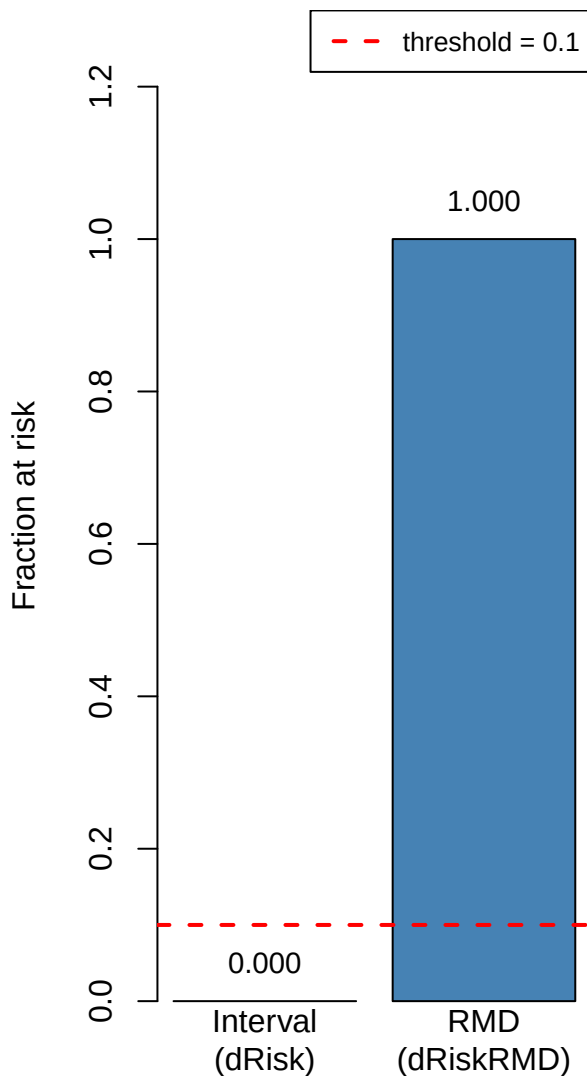


Minimum Robust Mahalanobis Distance

100.0% below threshold

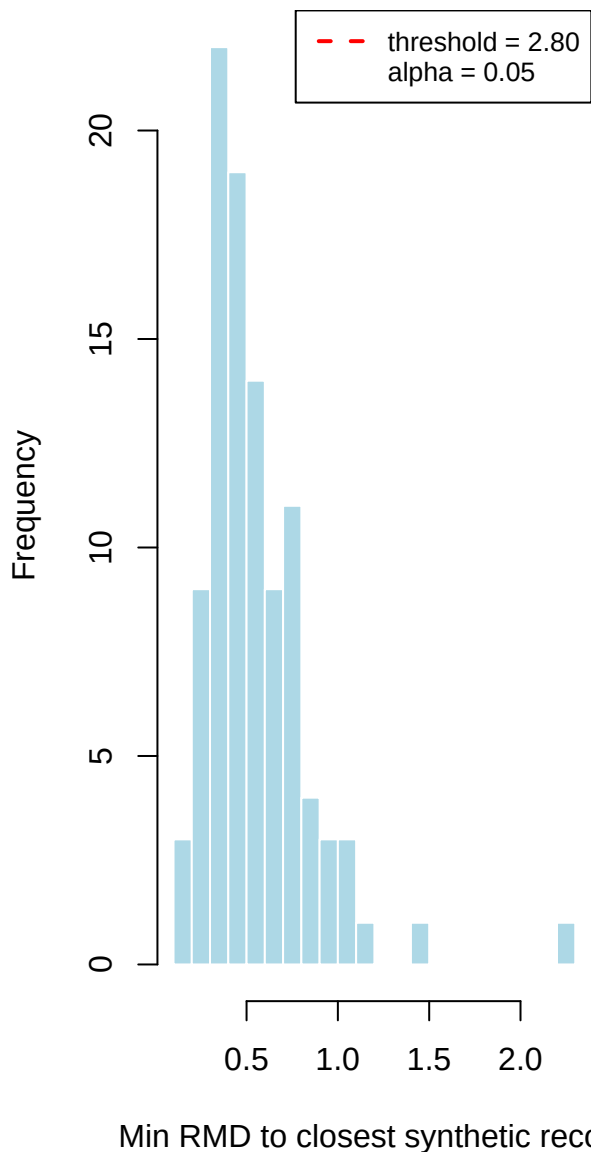


Disclosure Risk for Continuous Variables



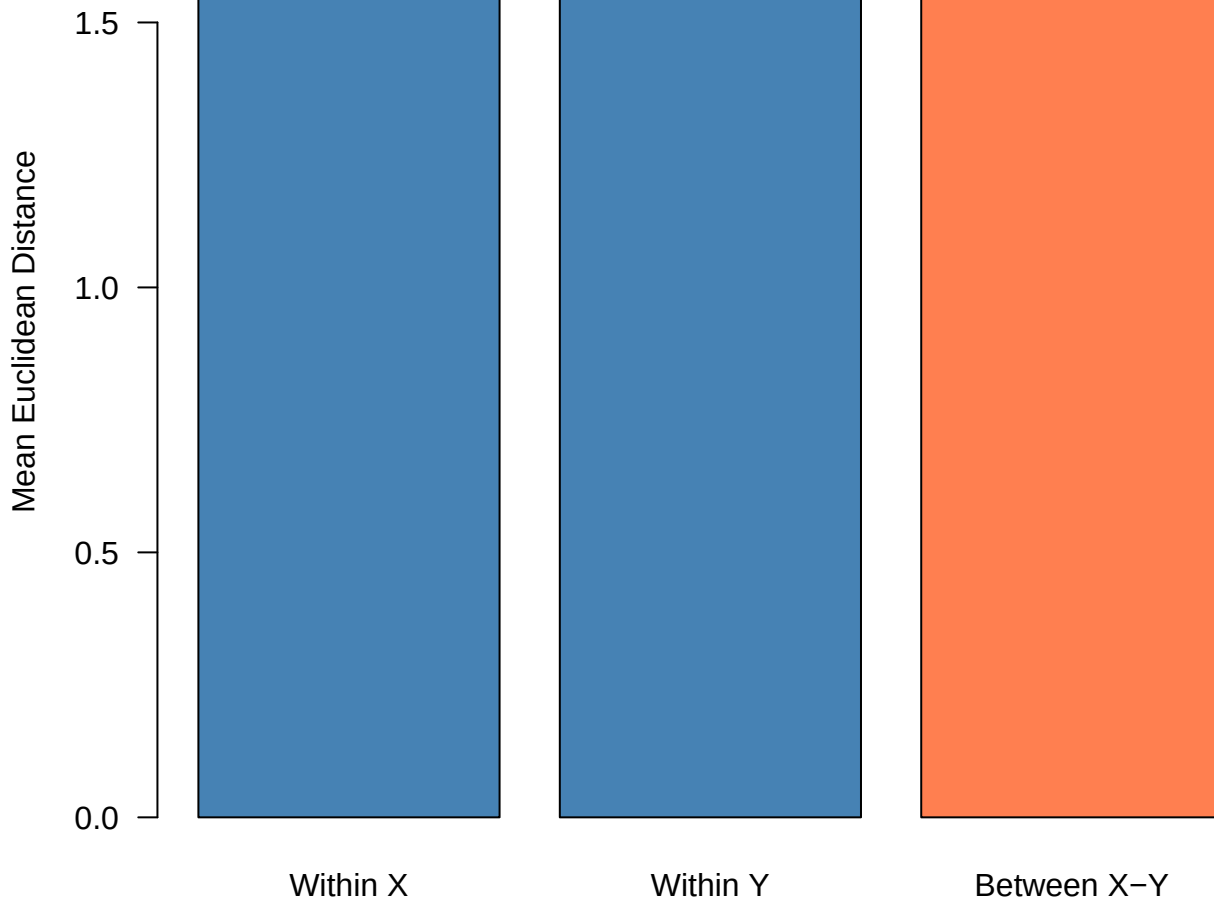
Minimum Robust Mahalanobis Distance

100.0% below threshold



Distance Components

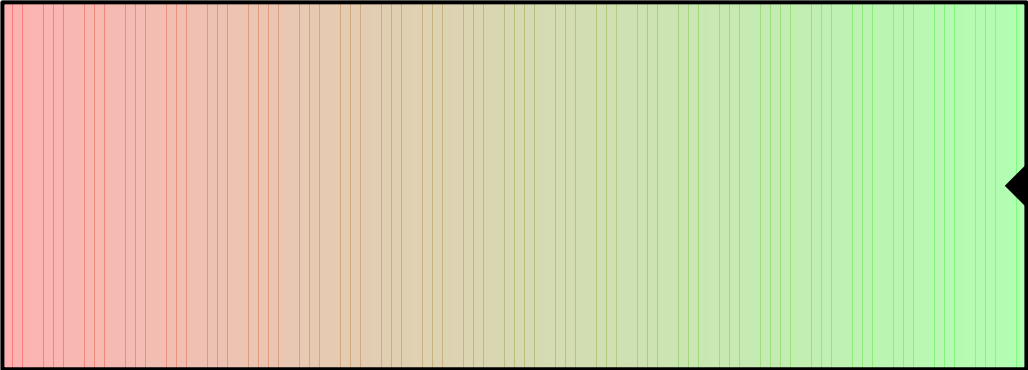
Energy Distance = 0.0000



Utility Score

Poor

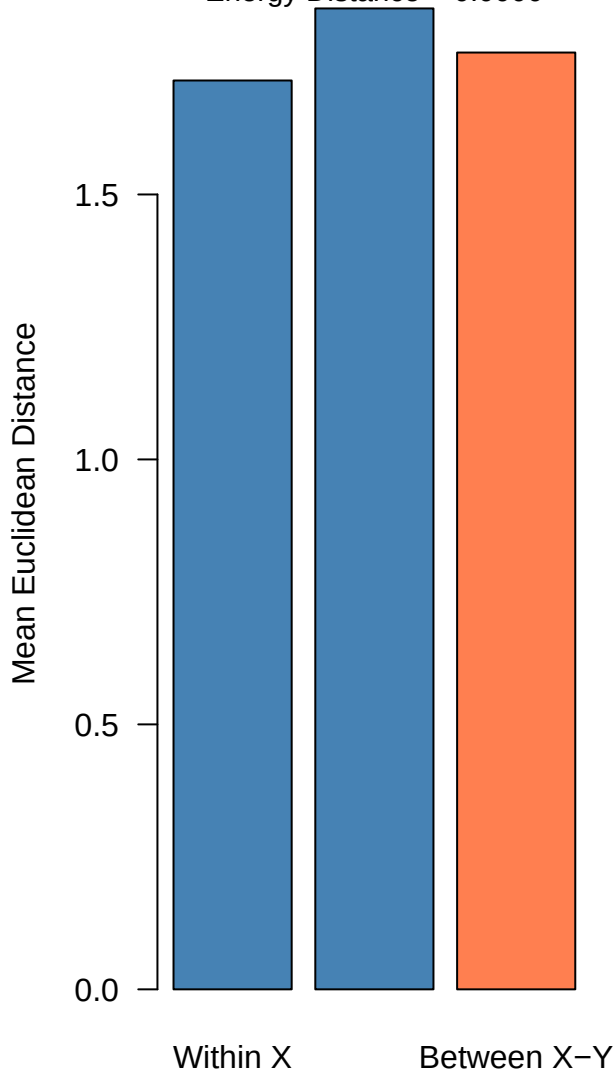
Good



Utility = 1.000

Distance Components

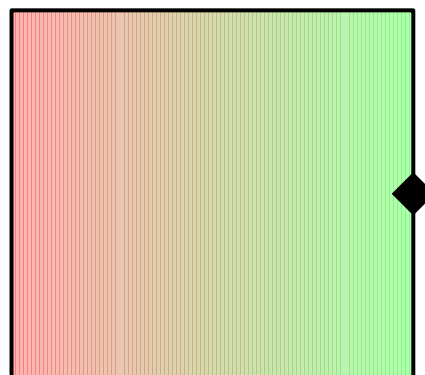
Energy Distance = 0.0000



Utility Score

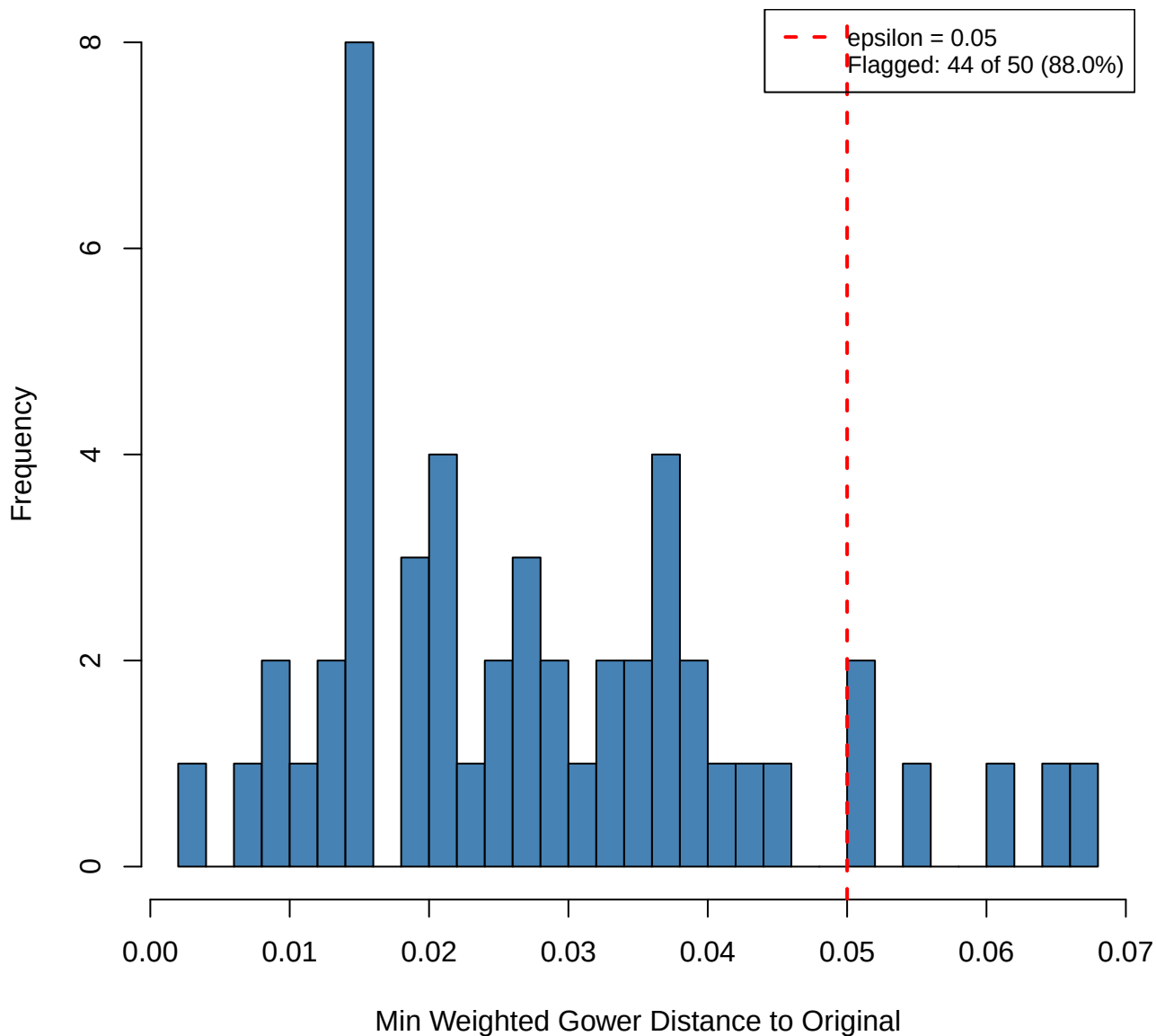
Poor

Good

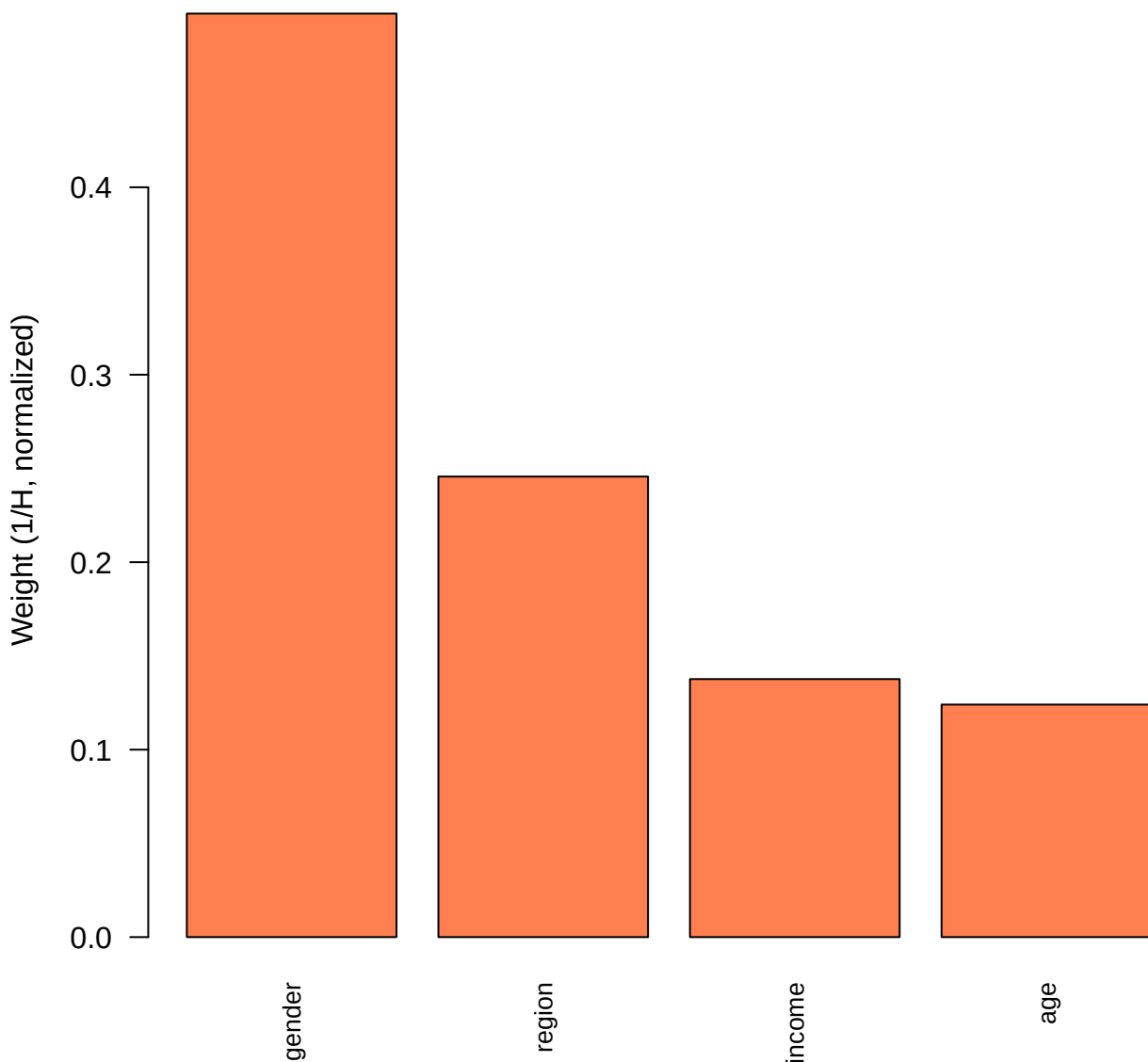


Utility = 1.000

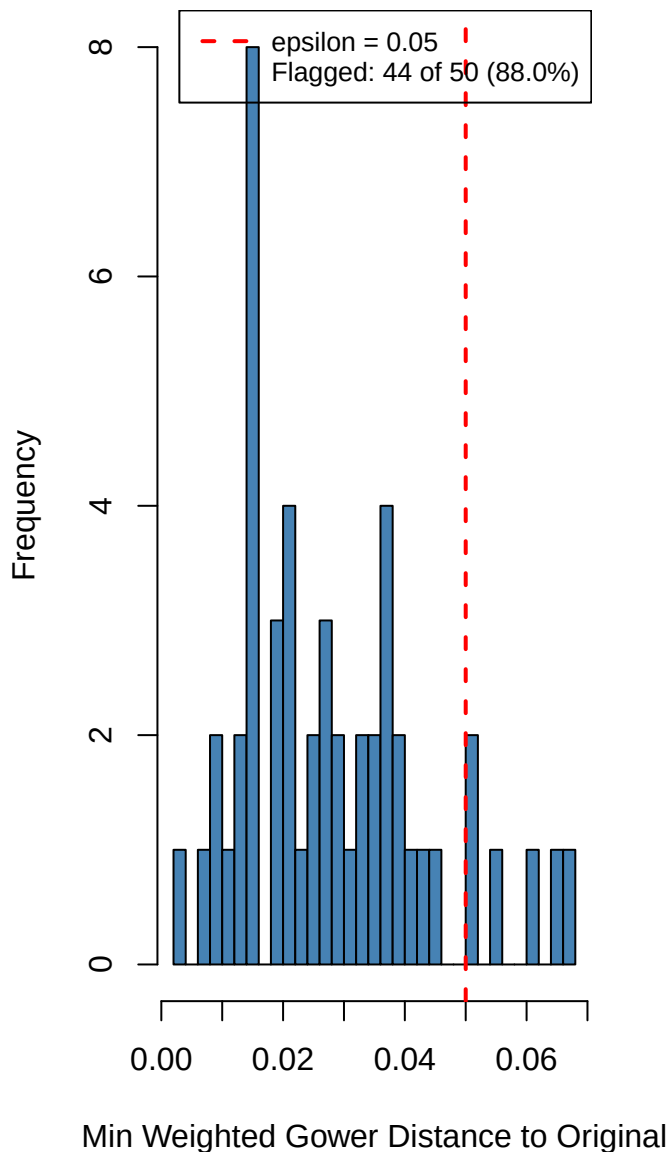
Minimum Weighted Distances (epsilon = 0.05)



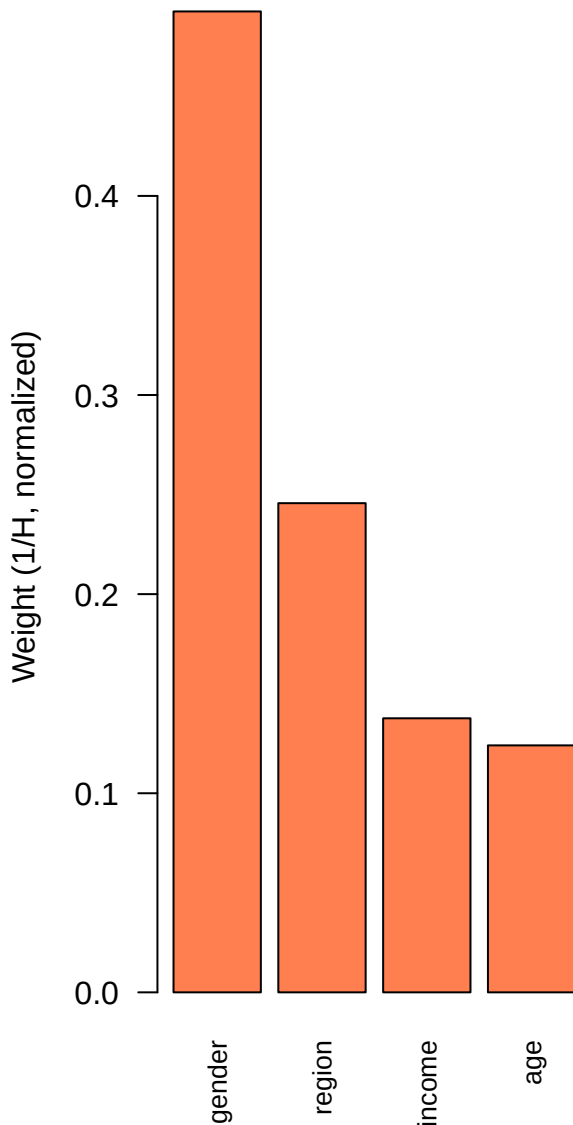
Entropy-Based Variable Weights



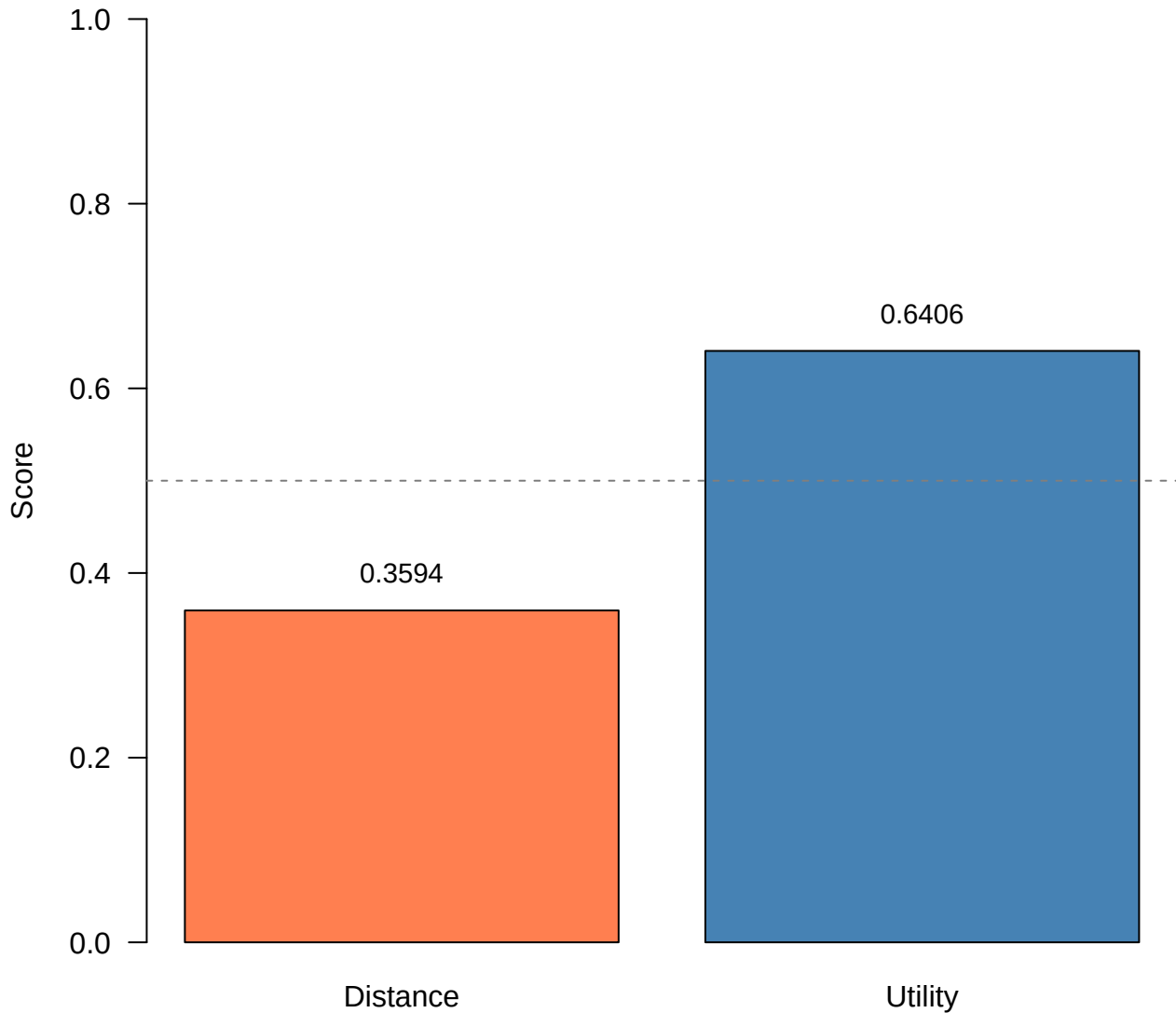
**Minimum Weighted Distances
(epsilon = 0.05)**



Entropy-Based Variable Weights

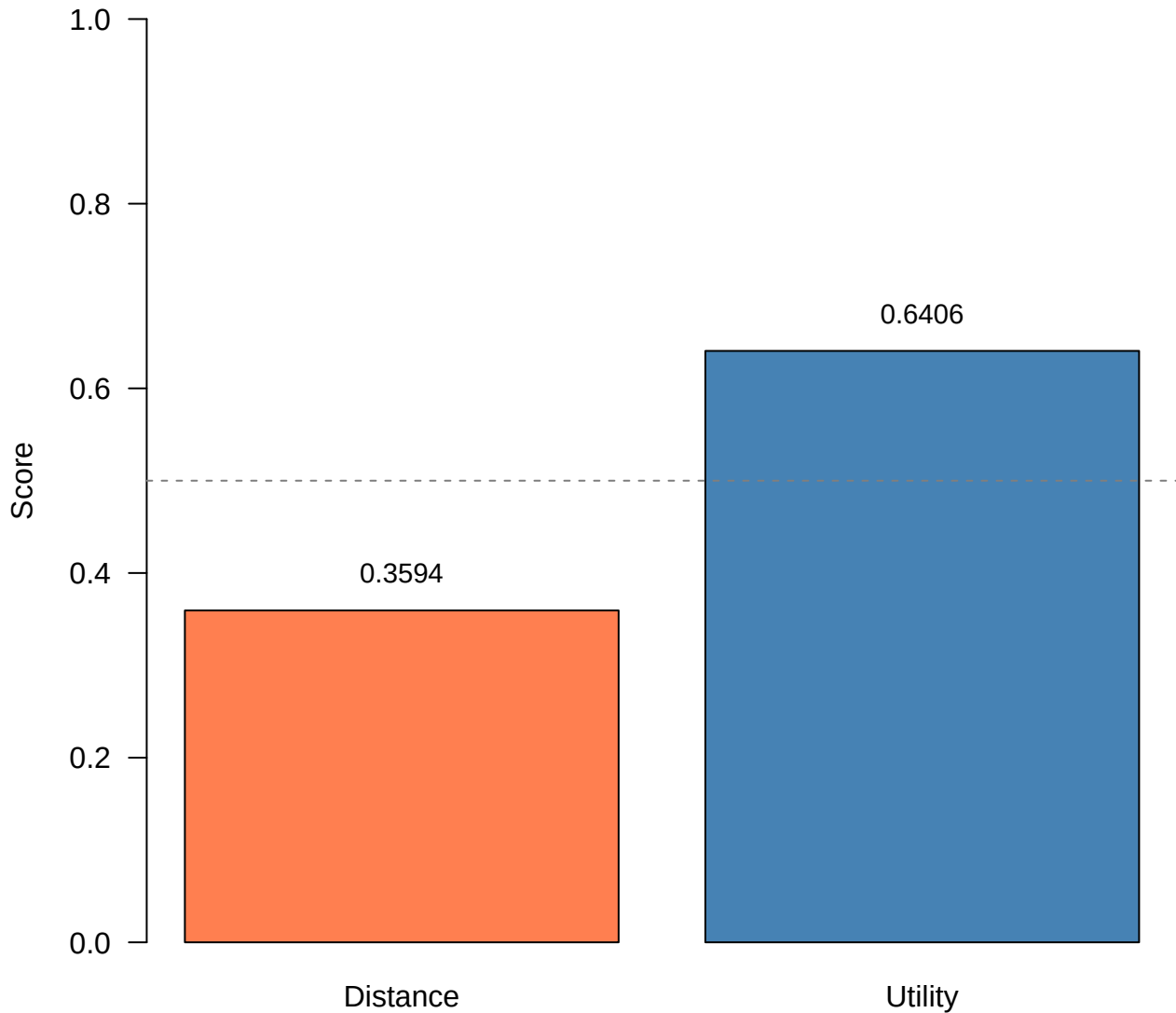


Average Gower Distance



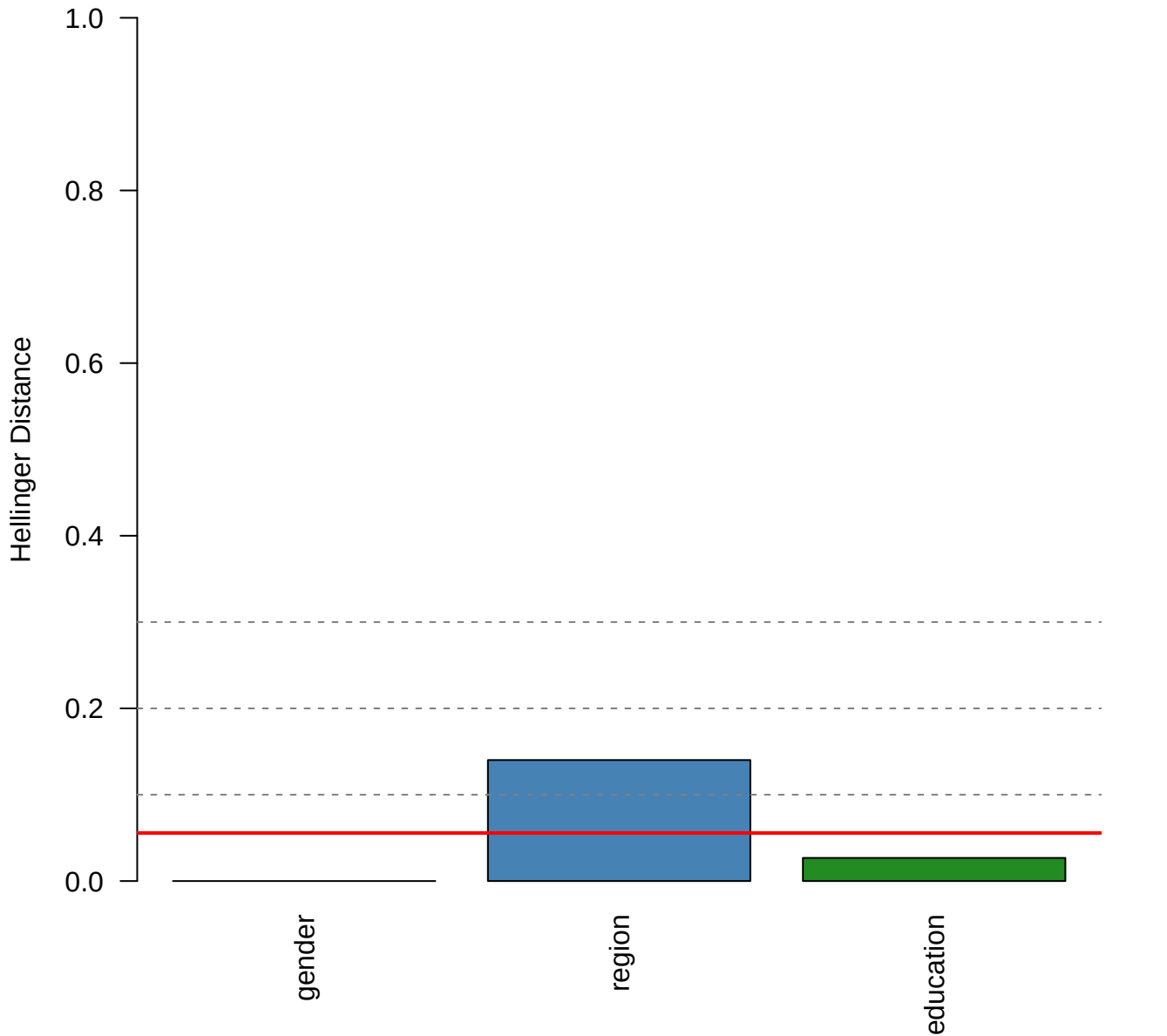
n = 20 obs, 4 vars

Average Gower Distance

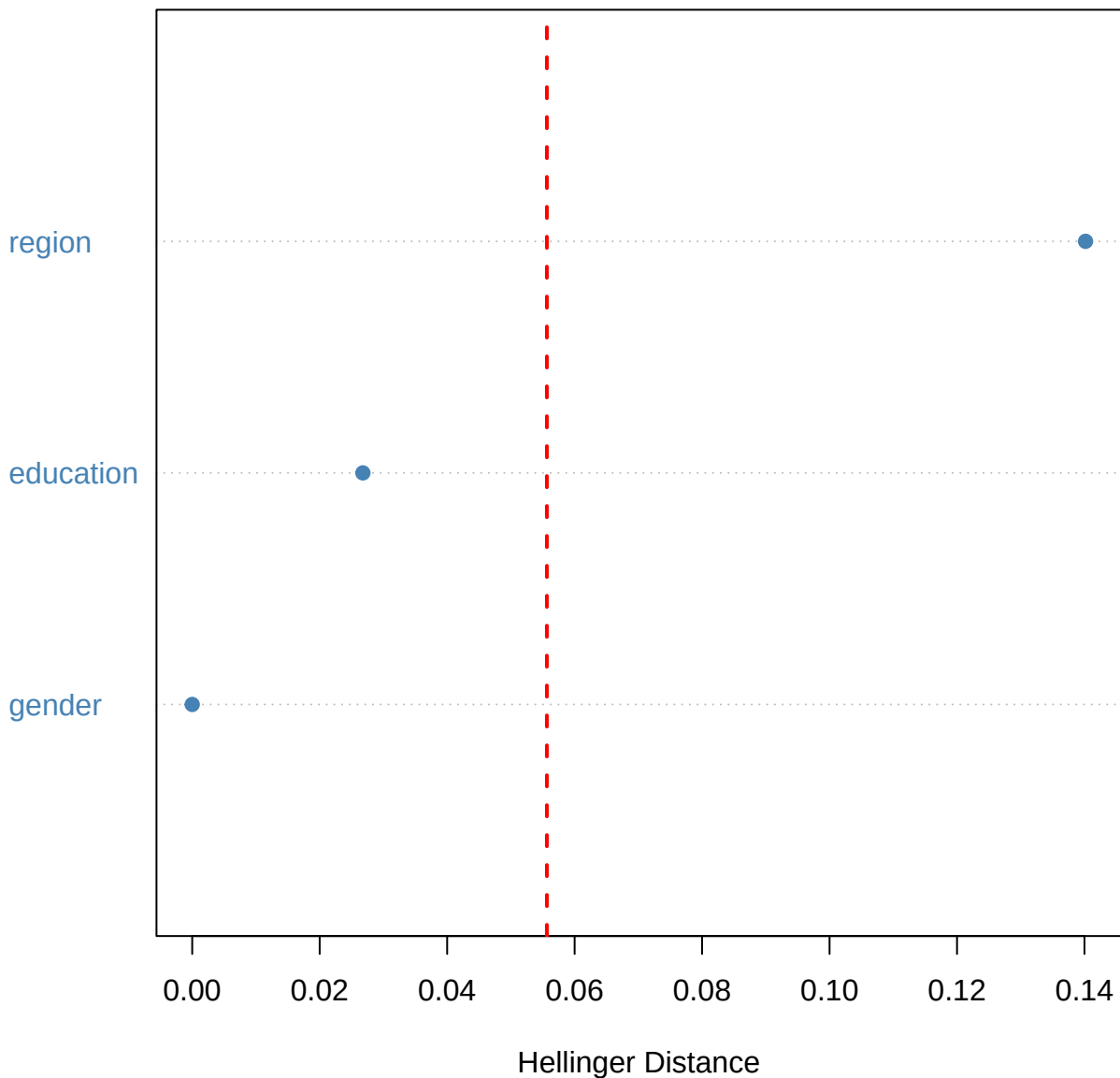


n = 20 obs, 4 vars

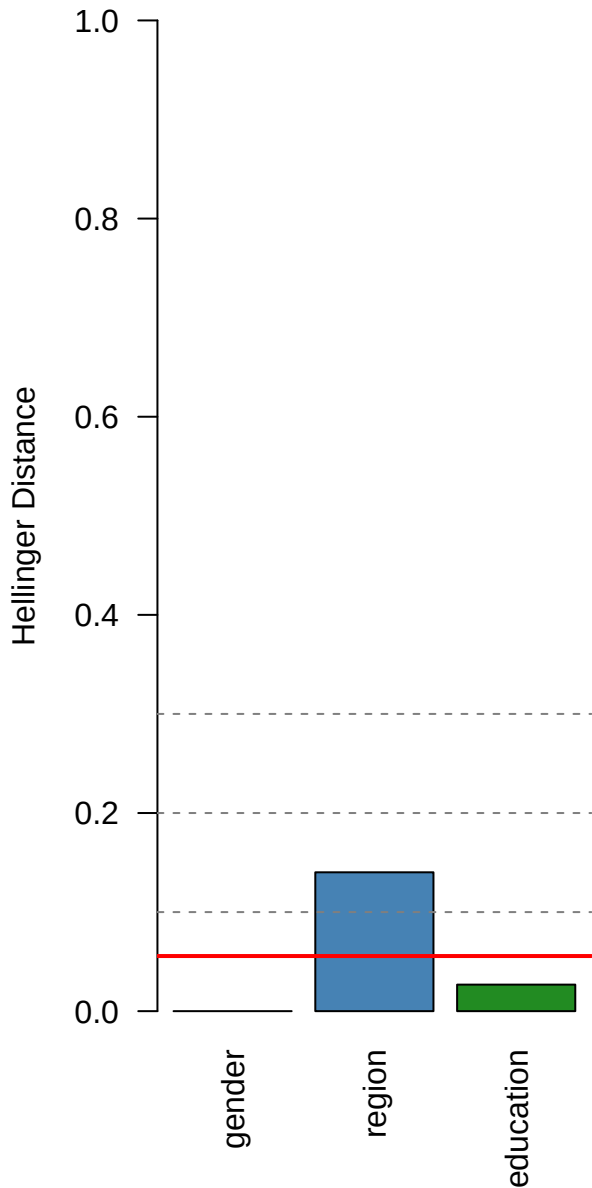
Hellinger Distance by Variable



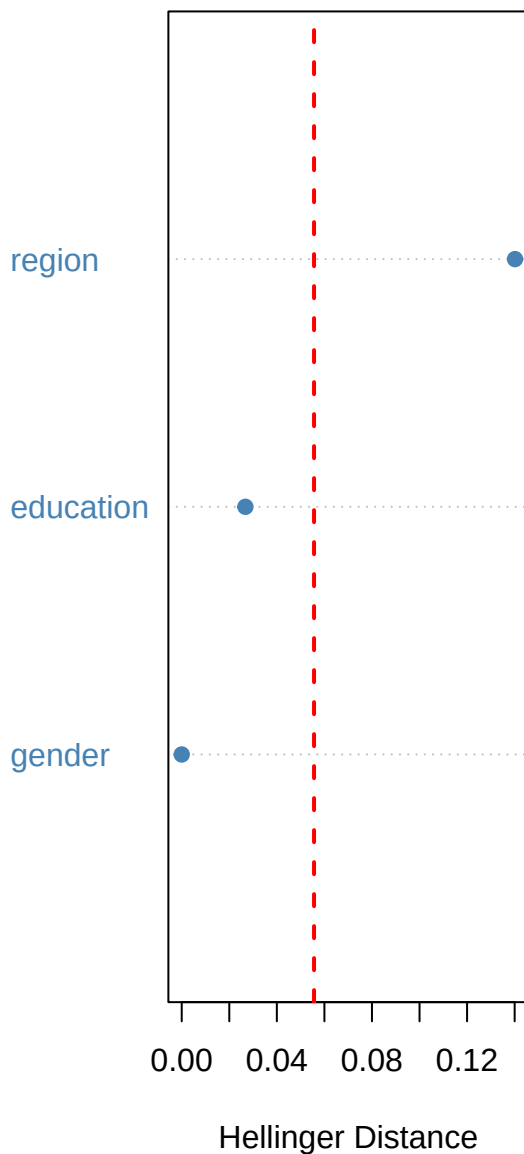
Hellinger Distance (sorted)



Hellinger Distance by Variable

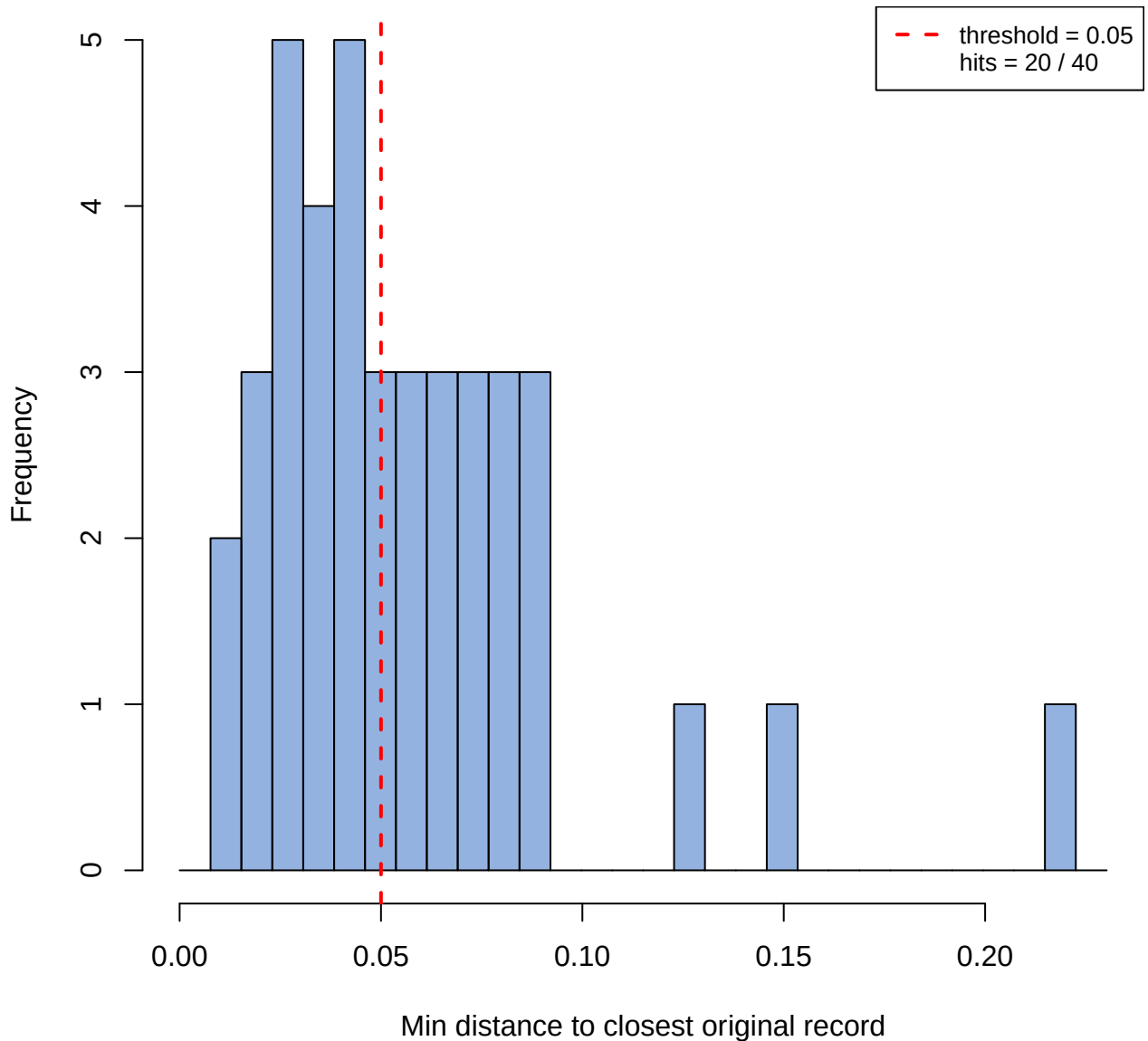


Hellinger Distance (sorted)

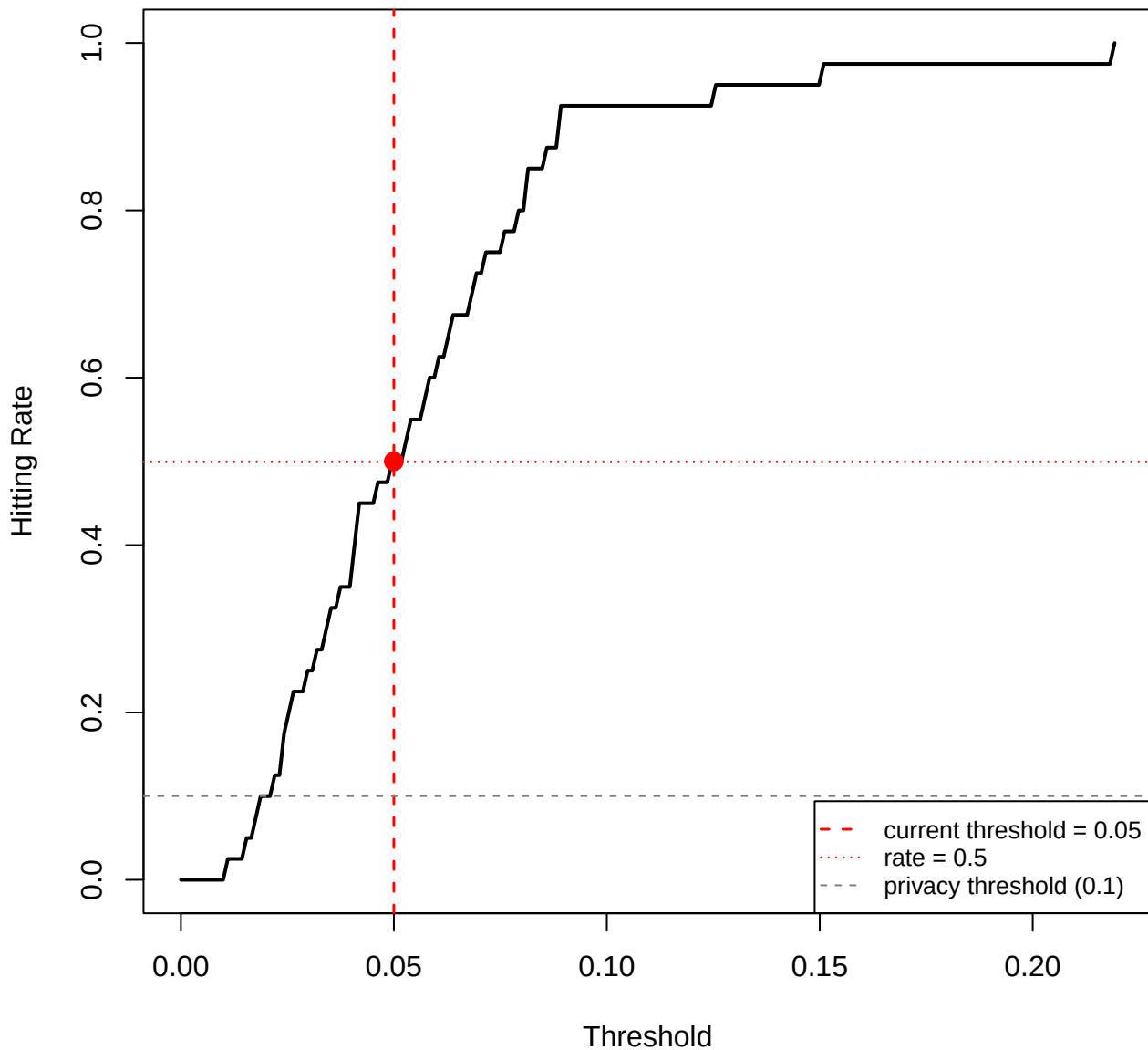


Min Distance Distribution

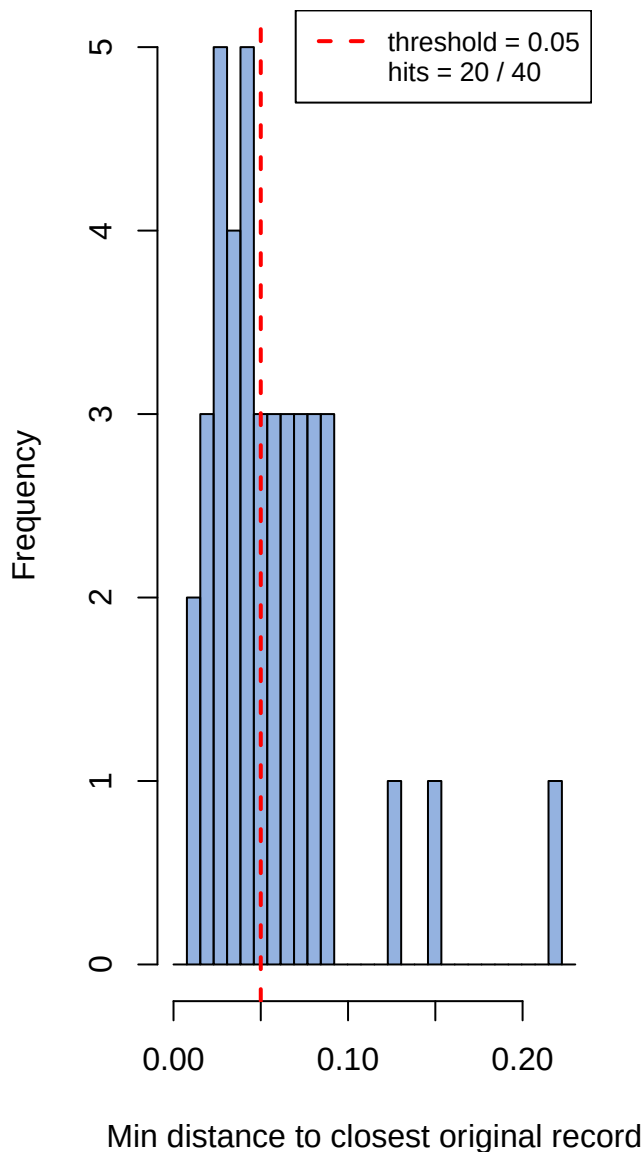
Hitting rate: 50.0%



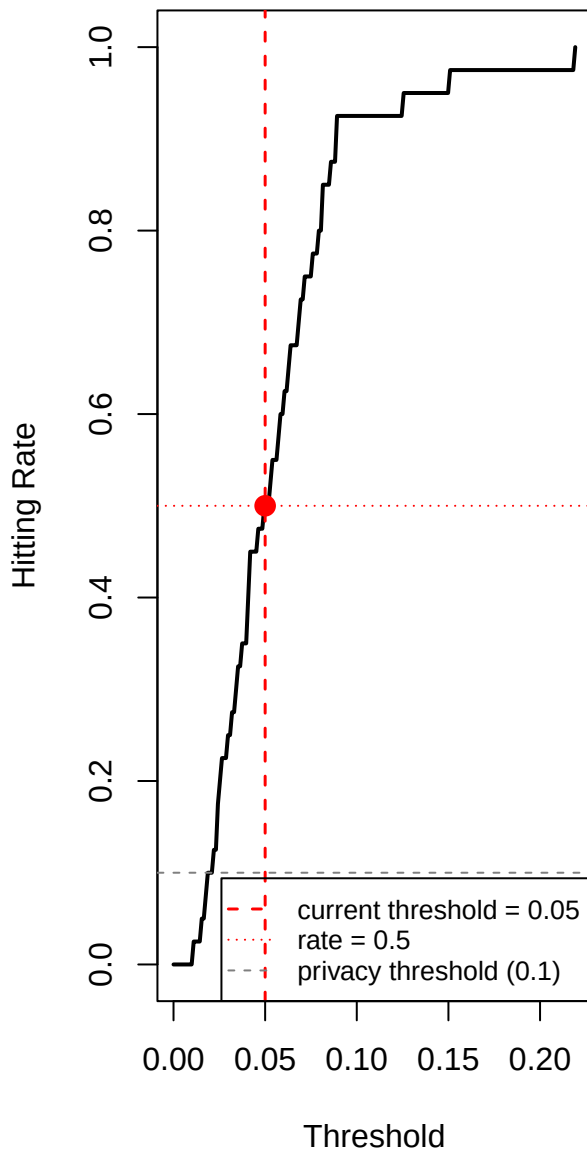
Hitting Rate vs Threshold



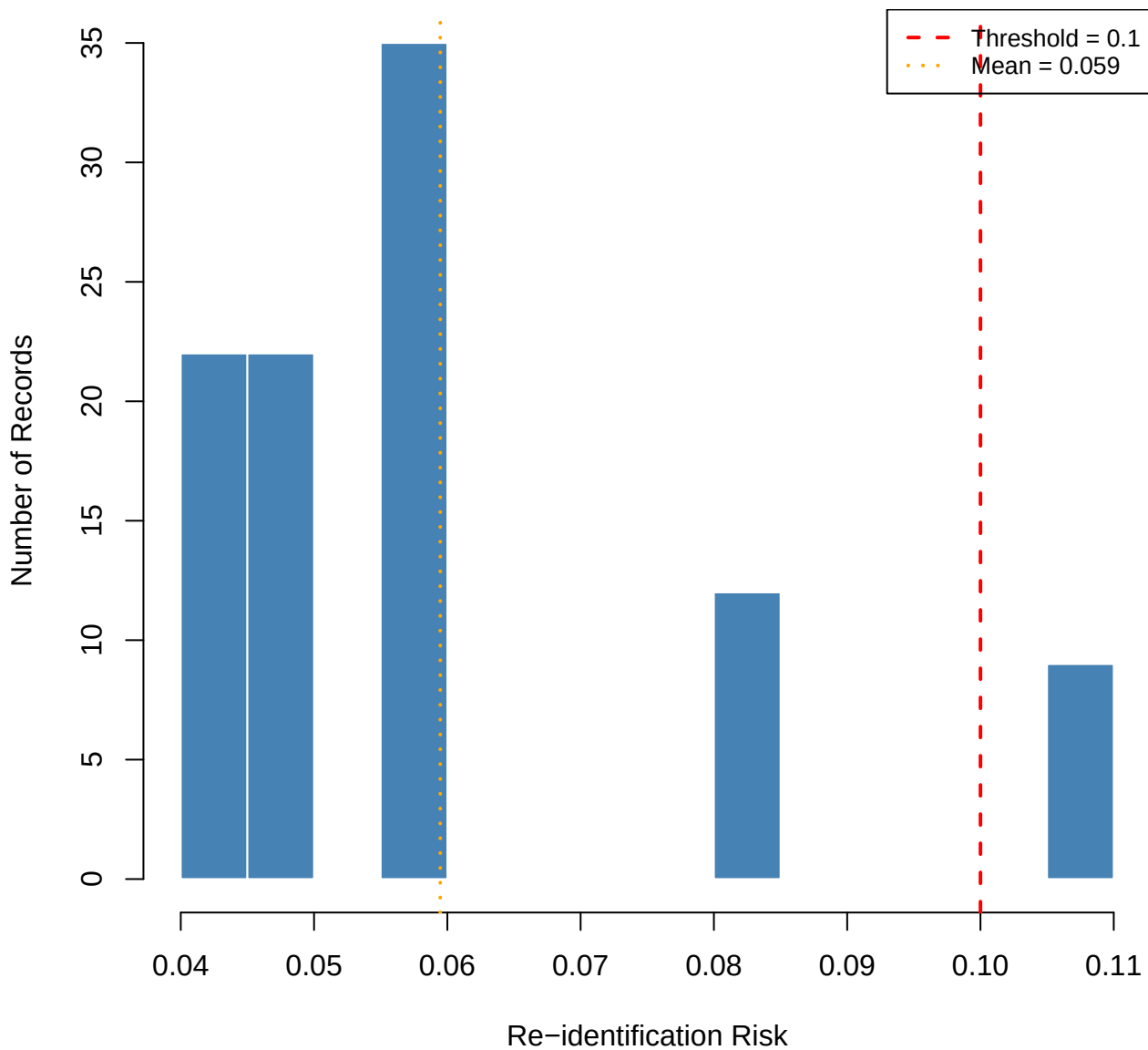
Min Distance Distribution
Hitting rate: 50.0%



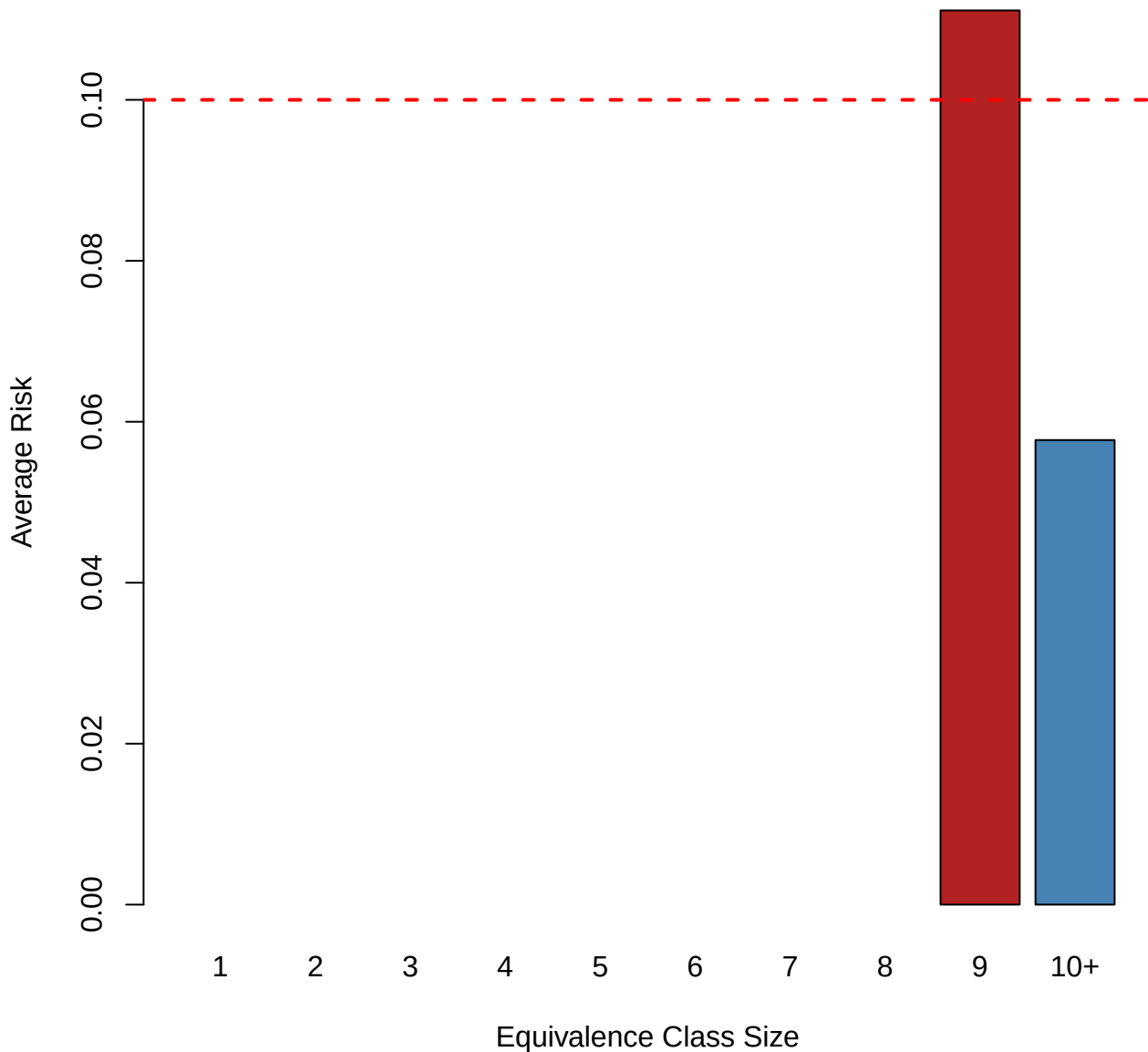
Hitting Rate vs Threshold



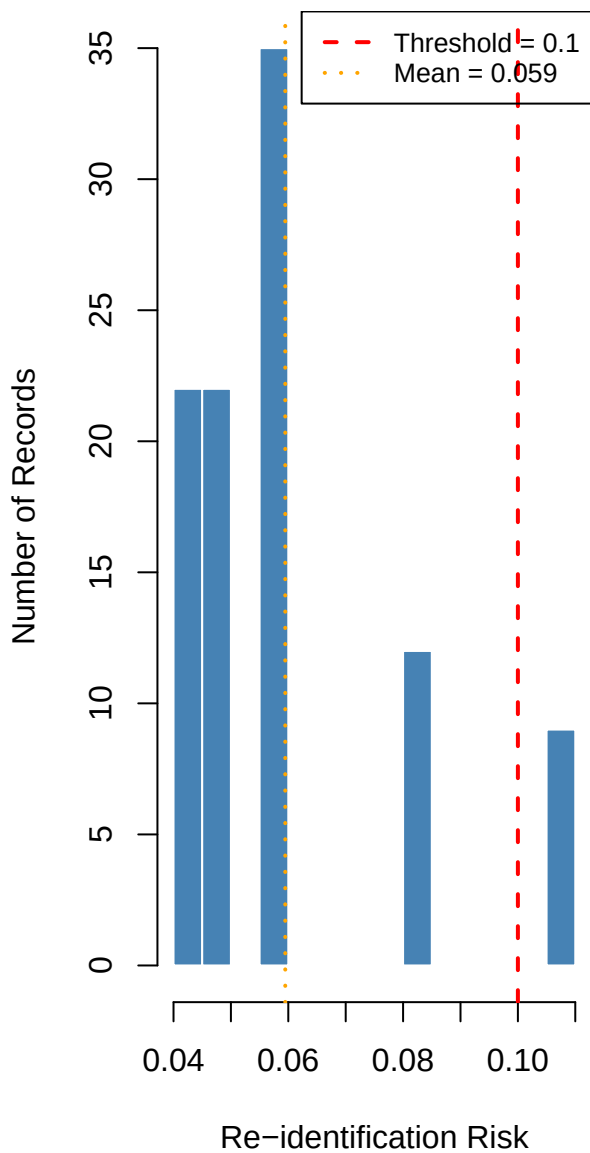
Distribution of Individual Risk



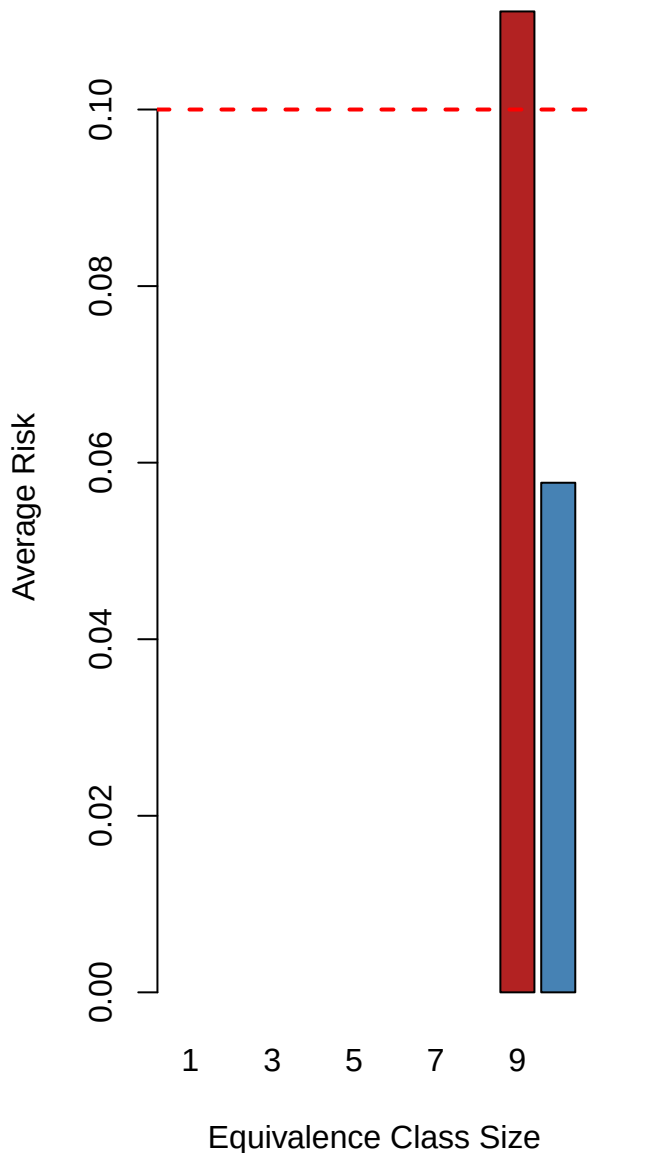
Average Risk by Equivalence Class Size



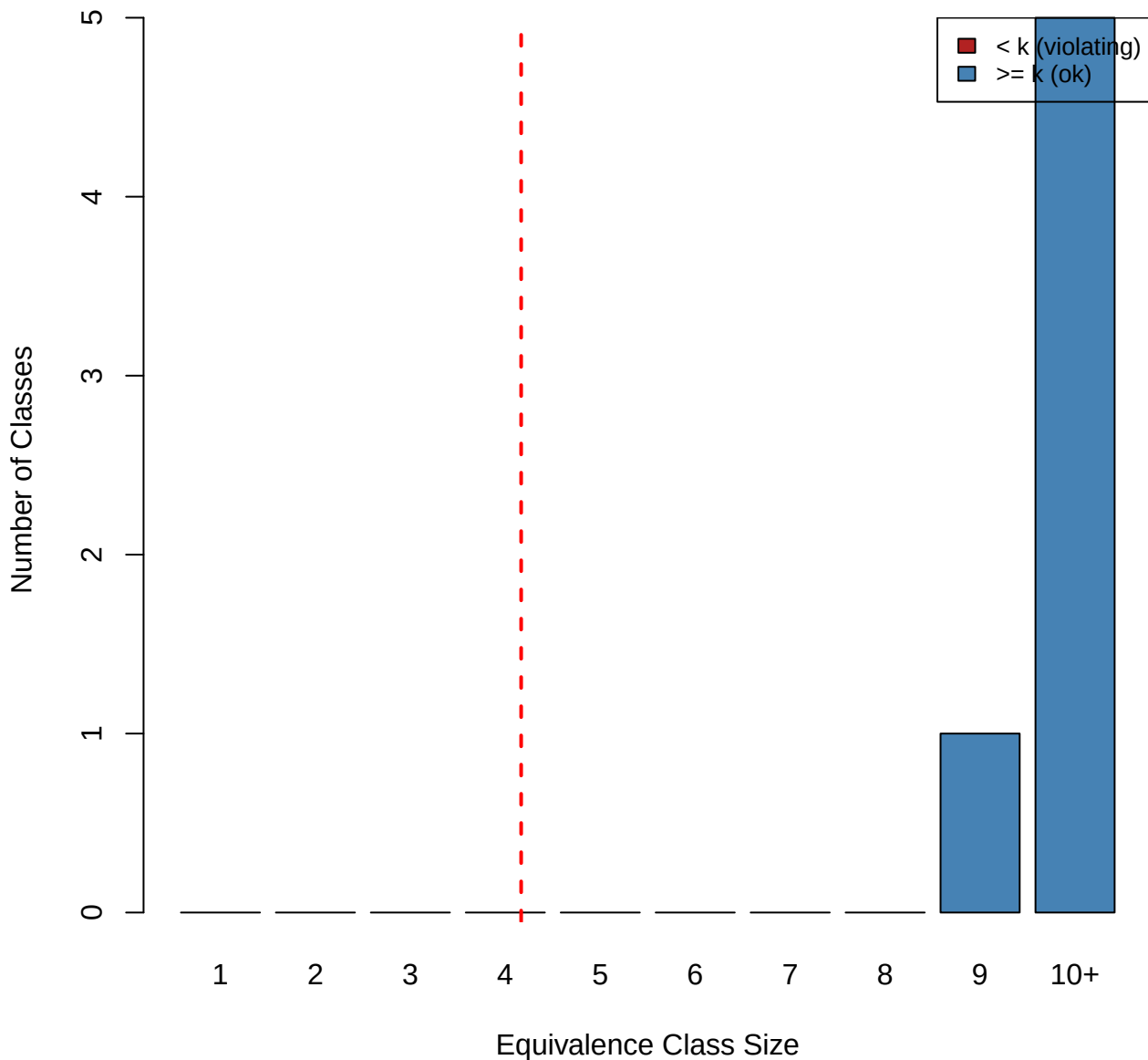
Distribution of Individual Risk



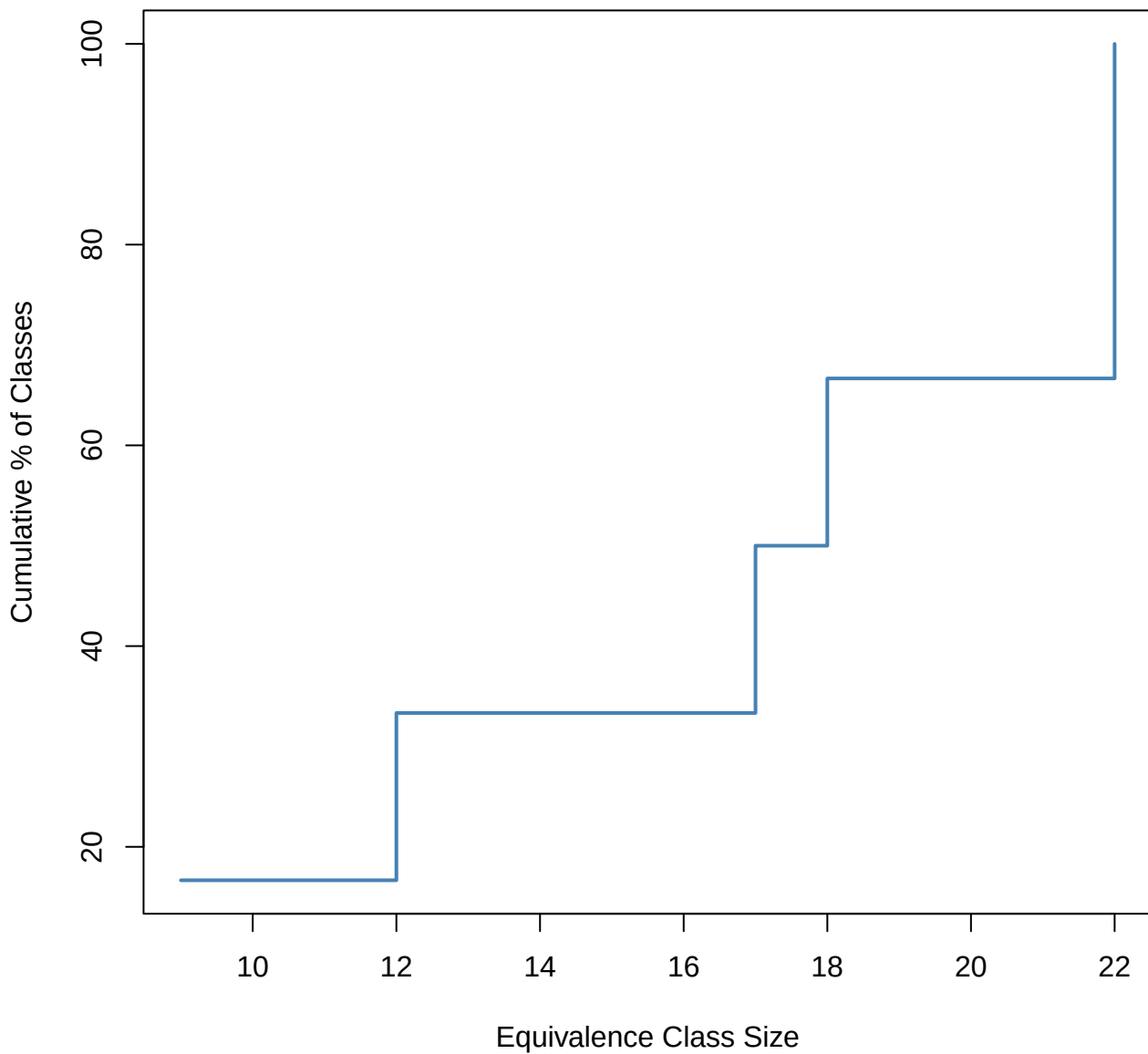
Average Risk by Equivalence Class Size



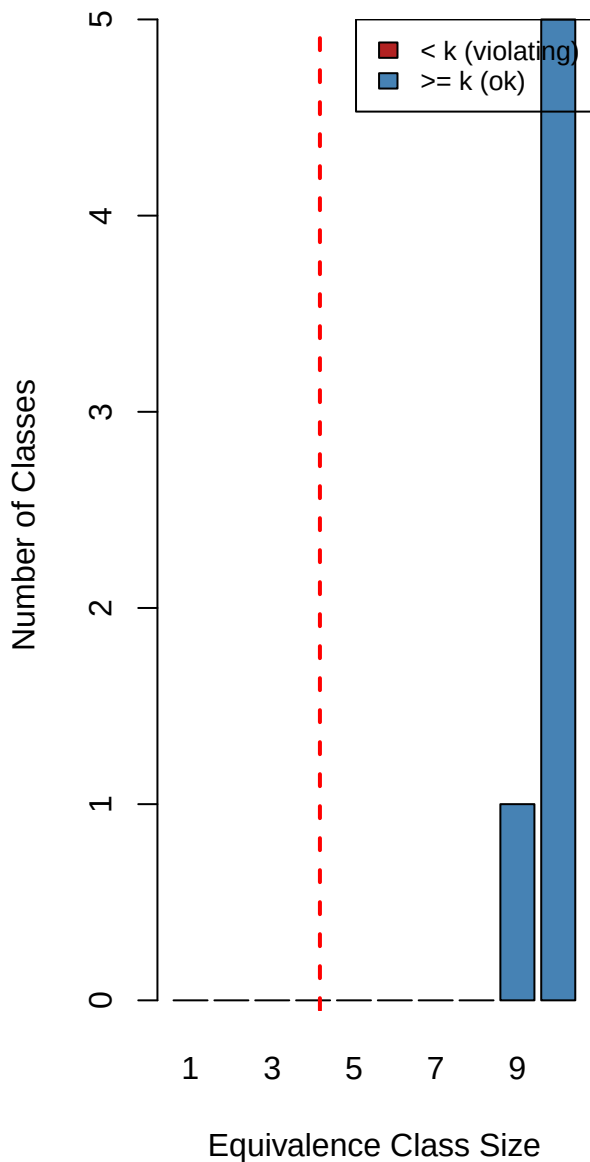
Equivalence Class Sizes (k = 5)



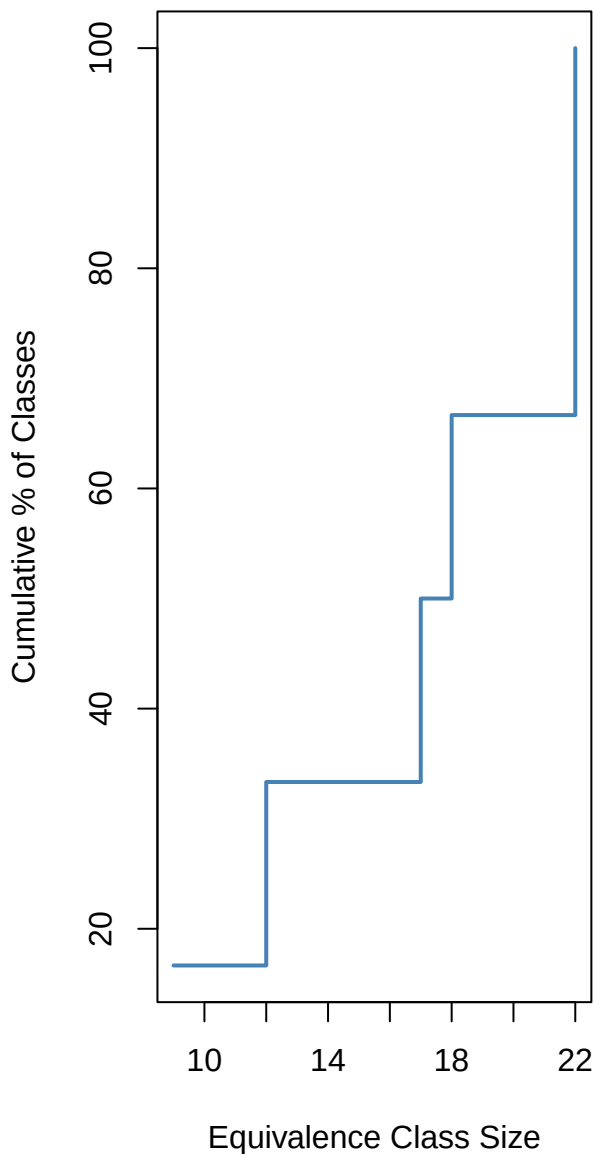
Cumulative Distribution of EC Sizes



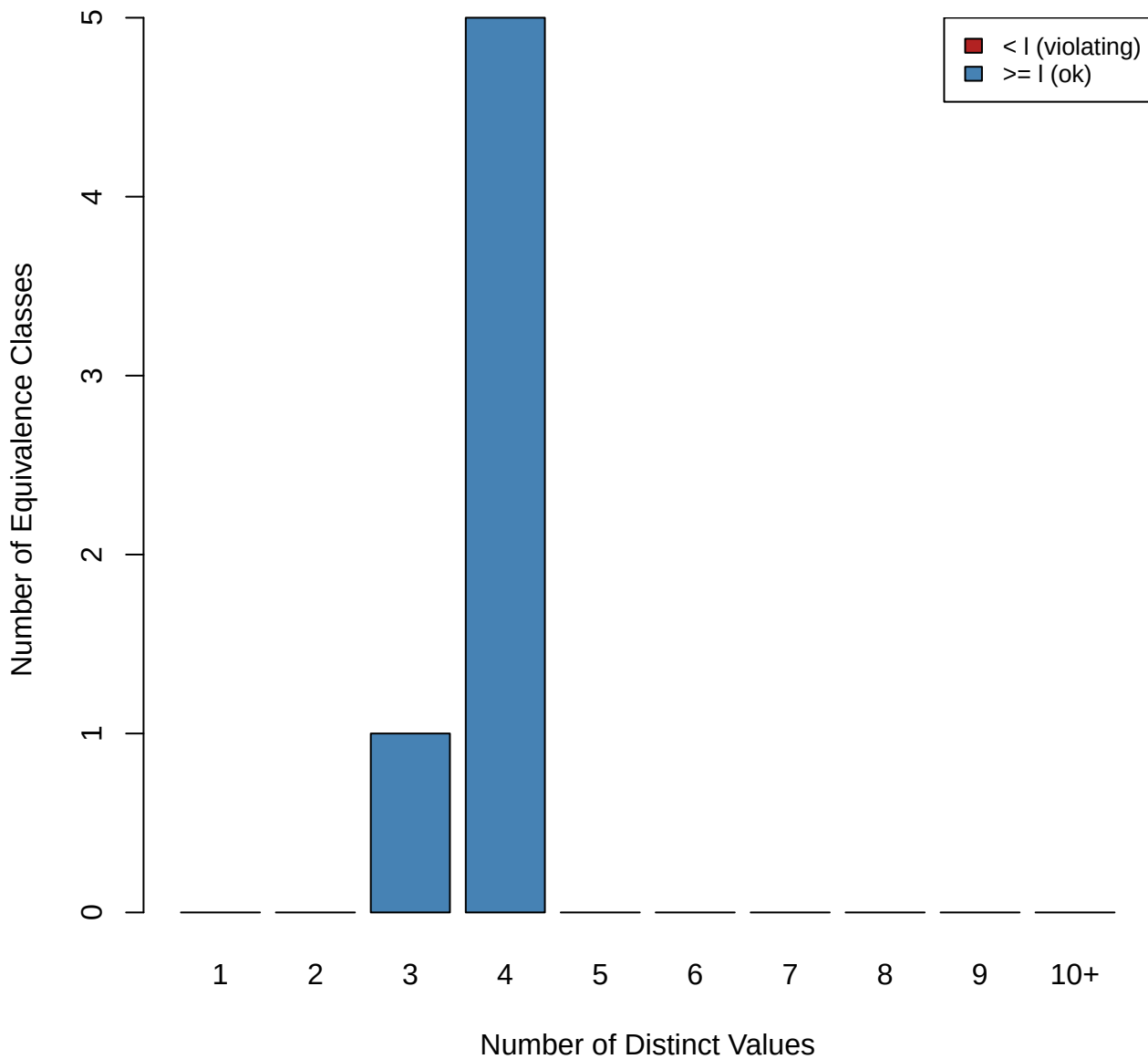
Equivalence Class Sizes (k = 5)



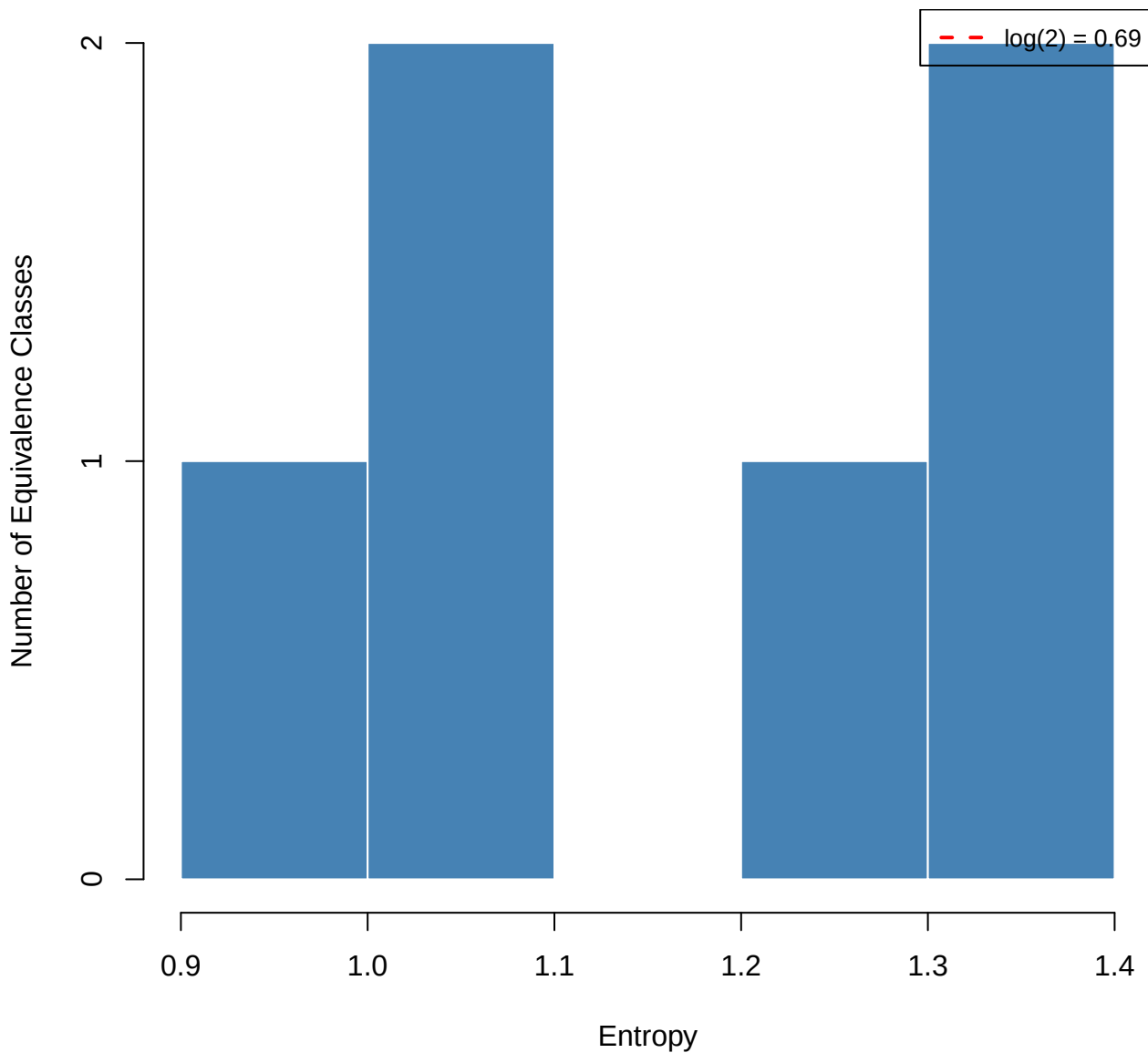
Cumulative Distribution of EC Size



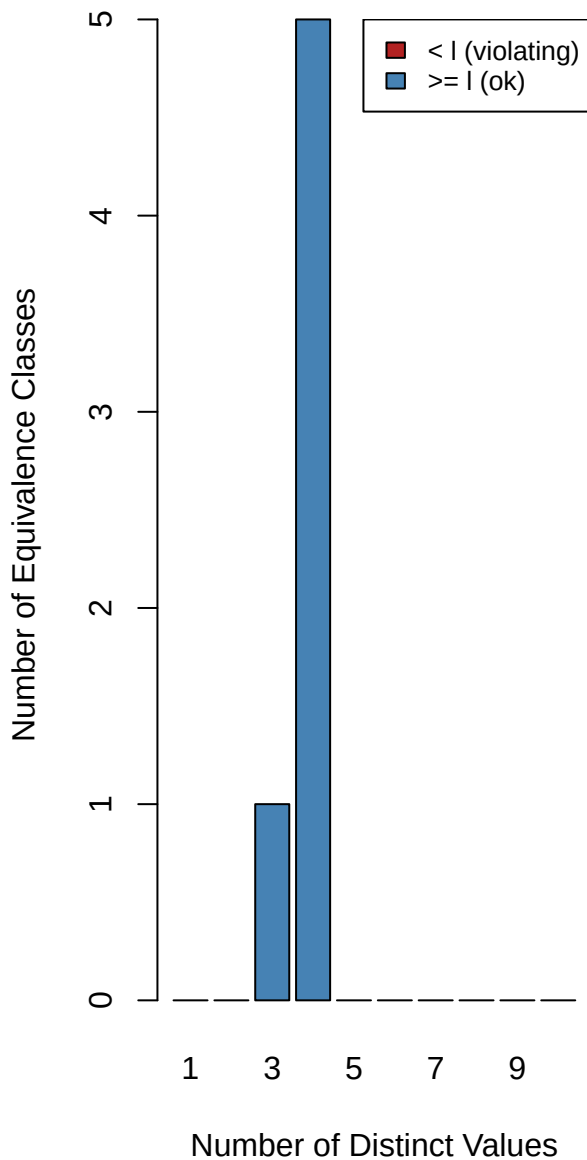
Distinct Sensitive Values per EC ($l = 2$)



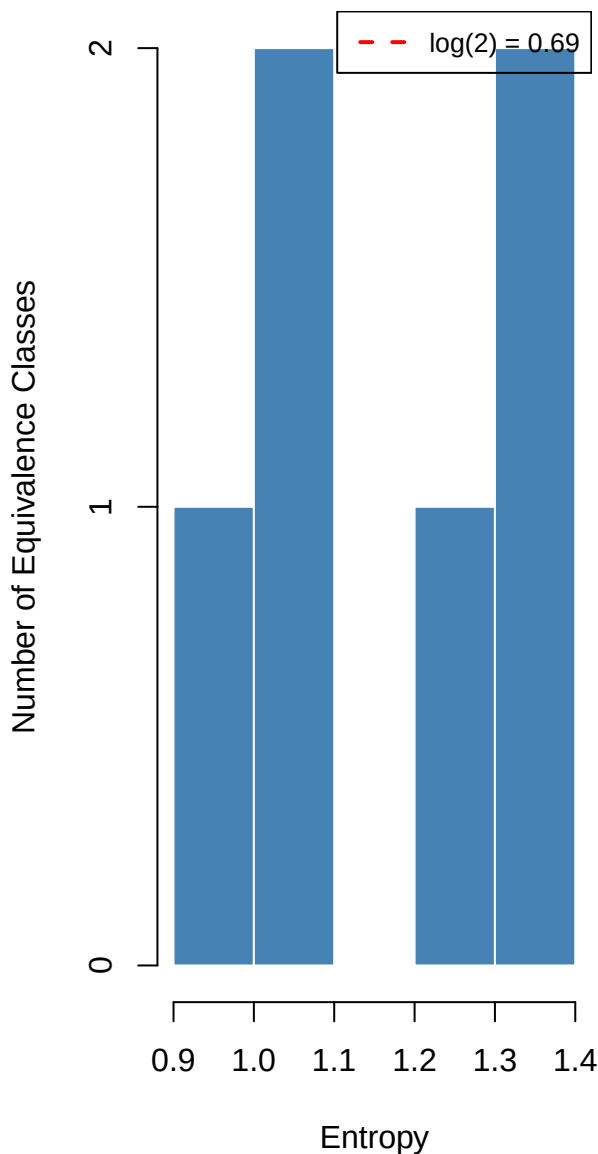
Entropy Distribution (l = 2)



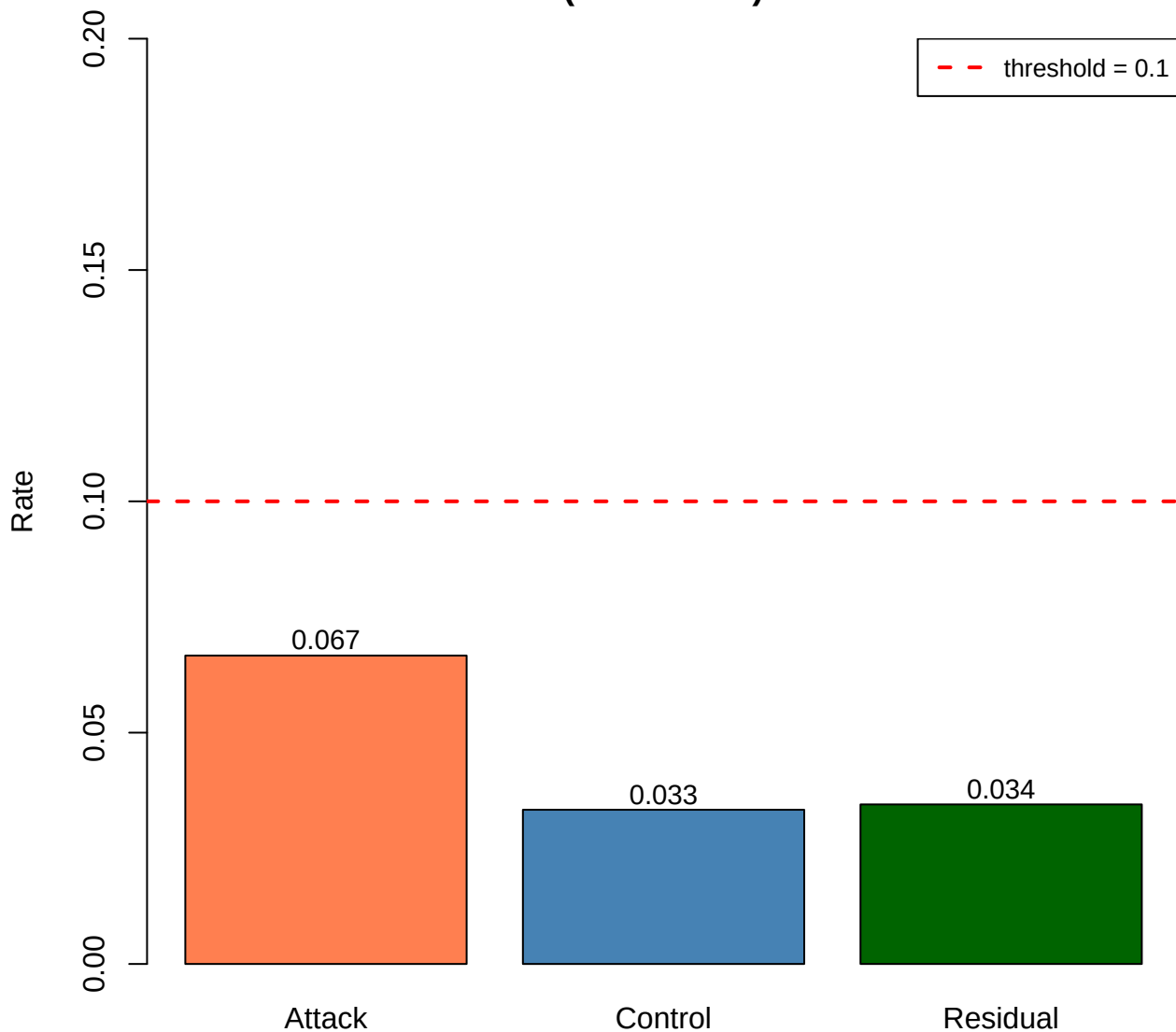
Distinct Sensitive Values per EC (l =



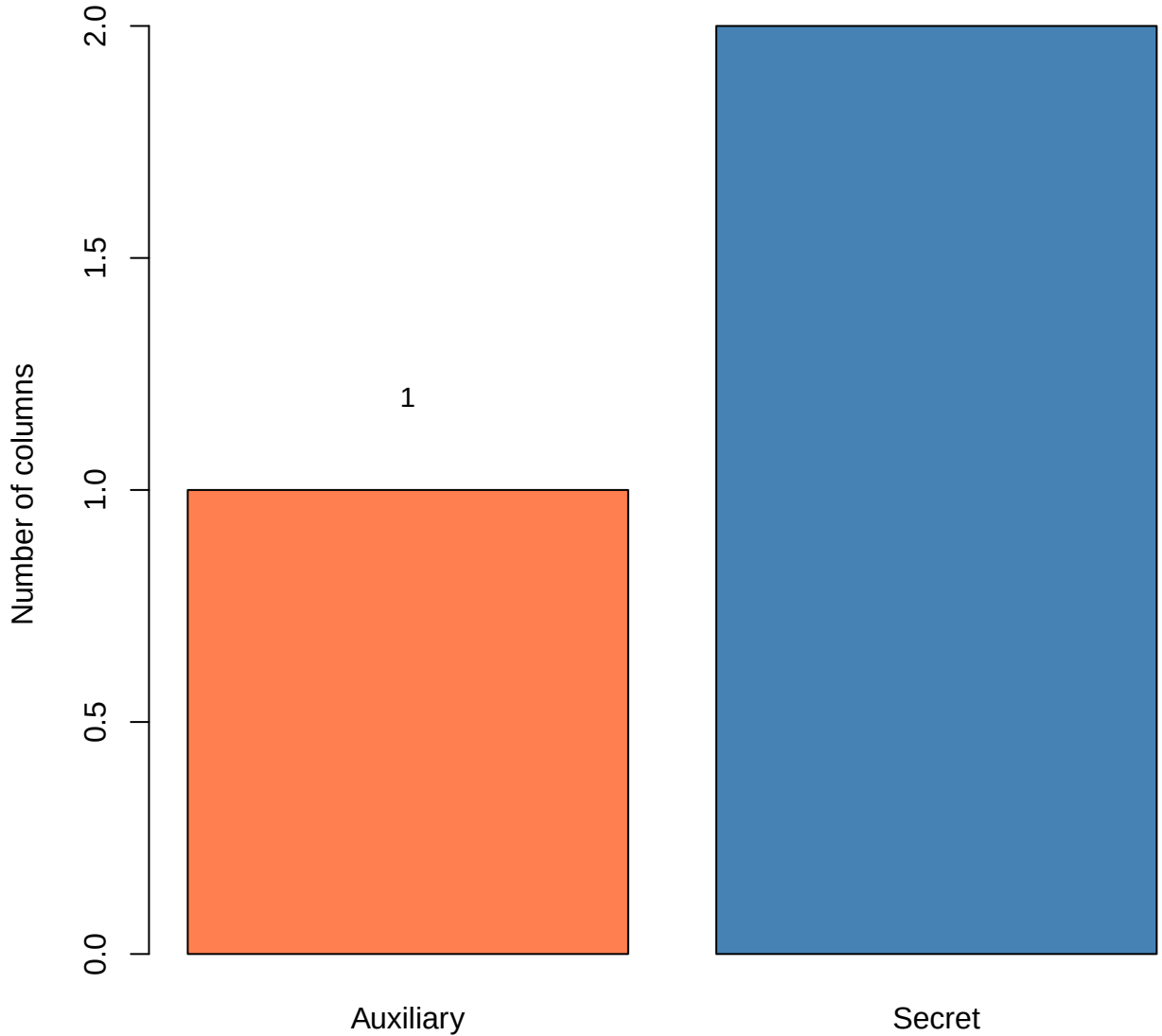
Entropy Distribution (l = 2)



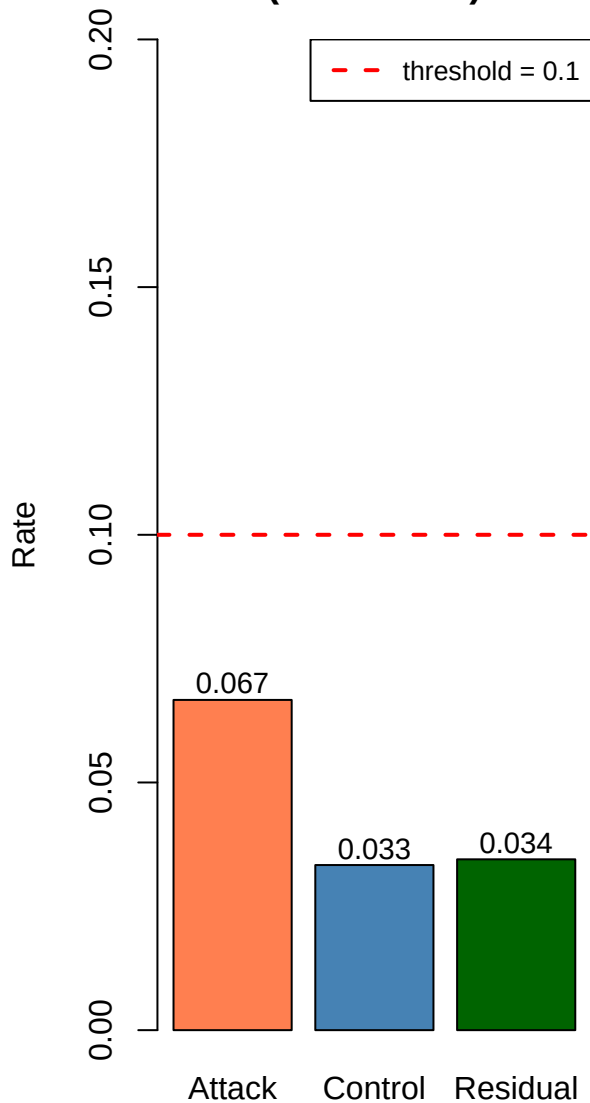
Linkability Risk (30 attacks)



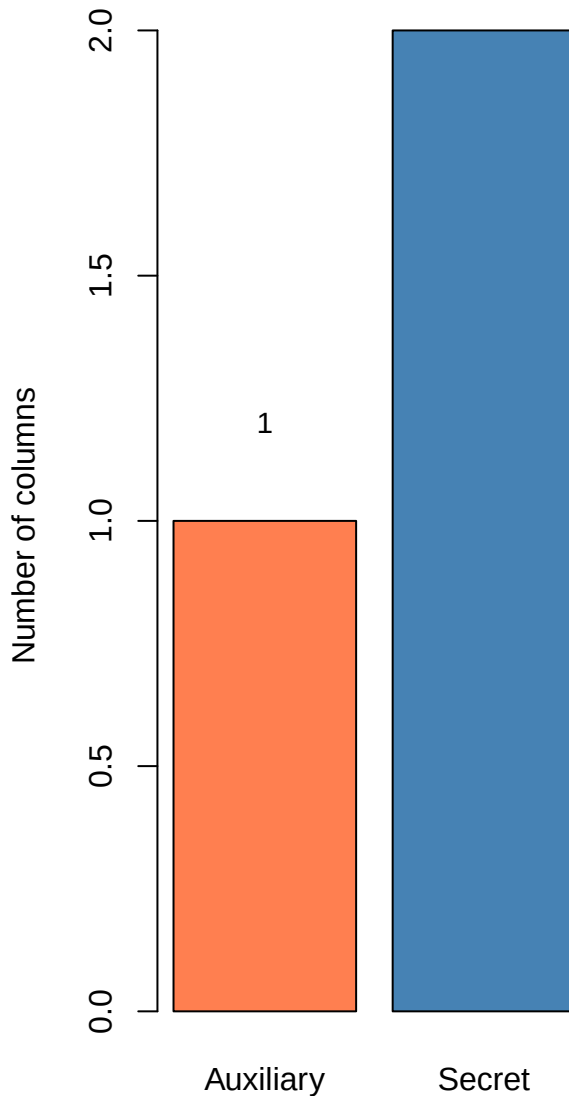
Column Split
1 aux + 2 secret



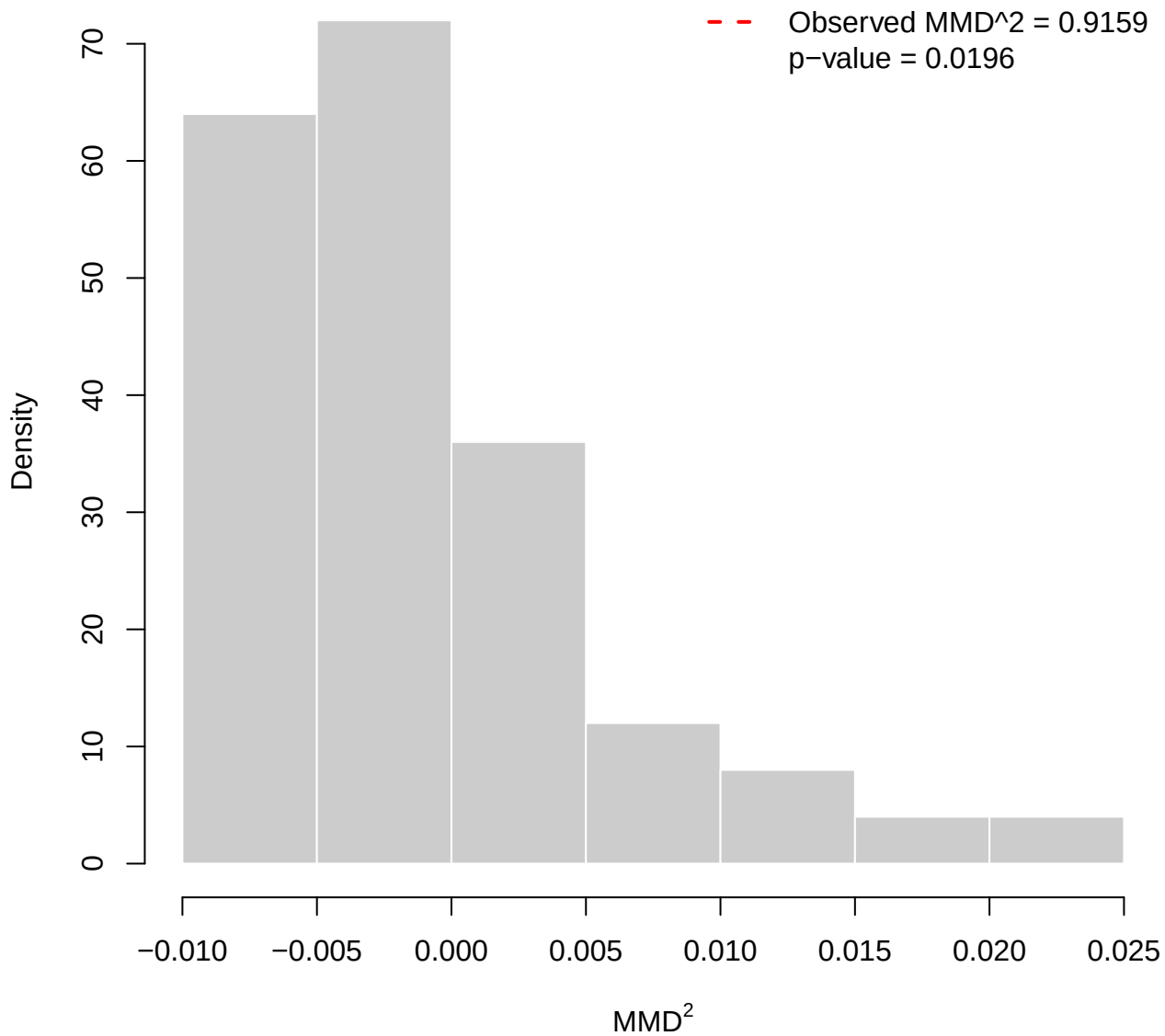
**Linkability Risk
(30 attacks)**



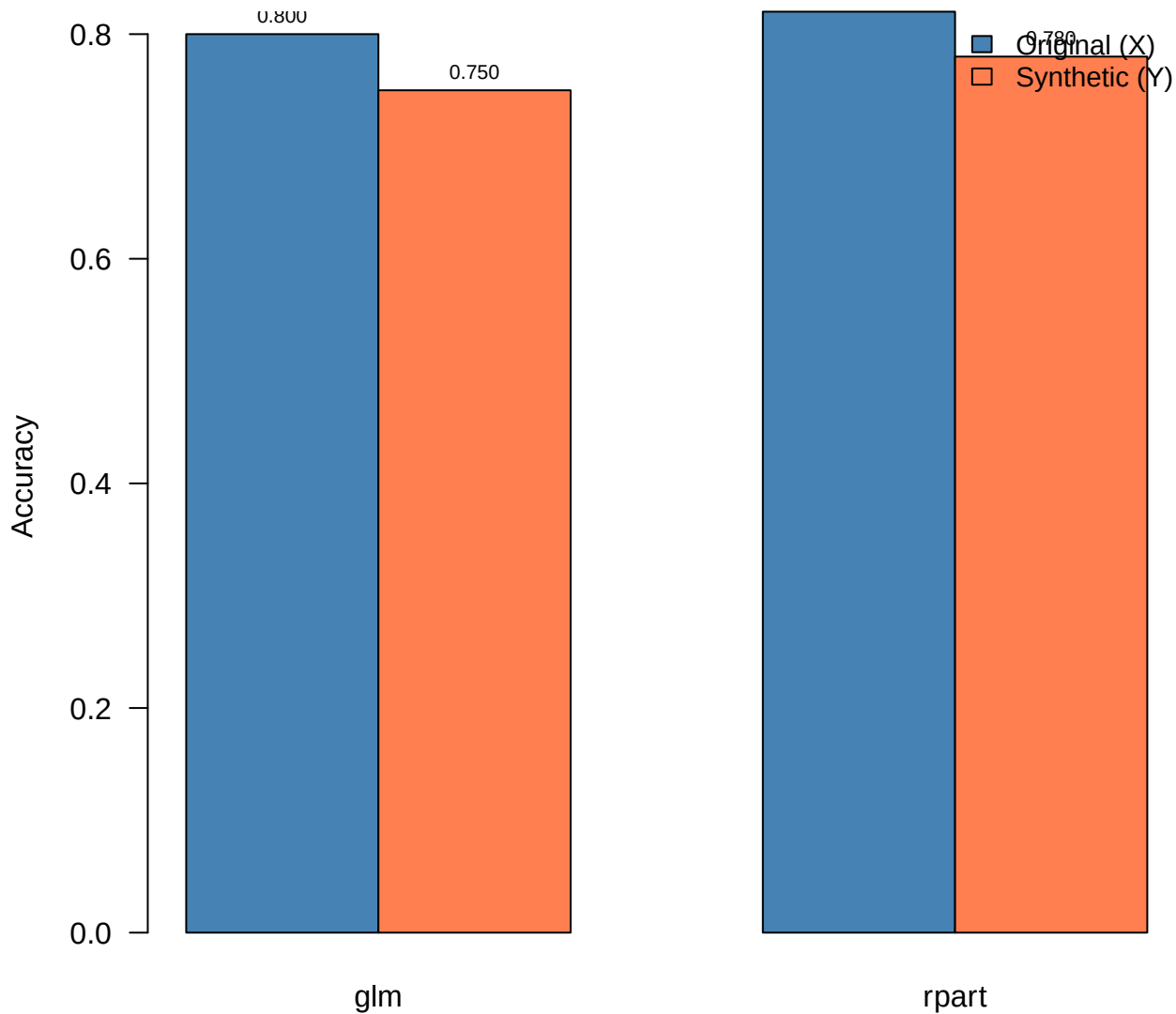
**Column Split
1 aux + 2 secret**



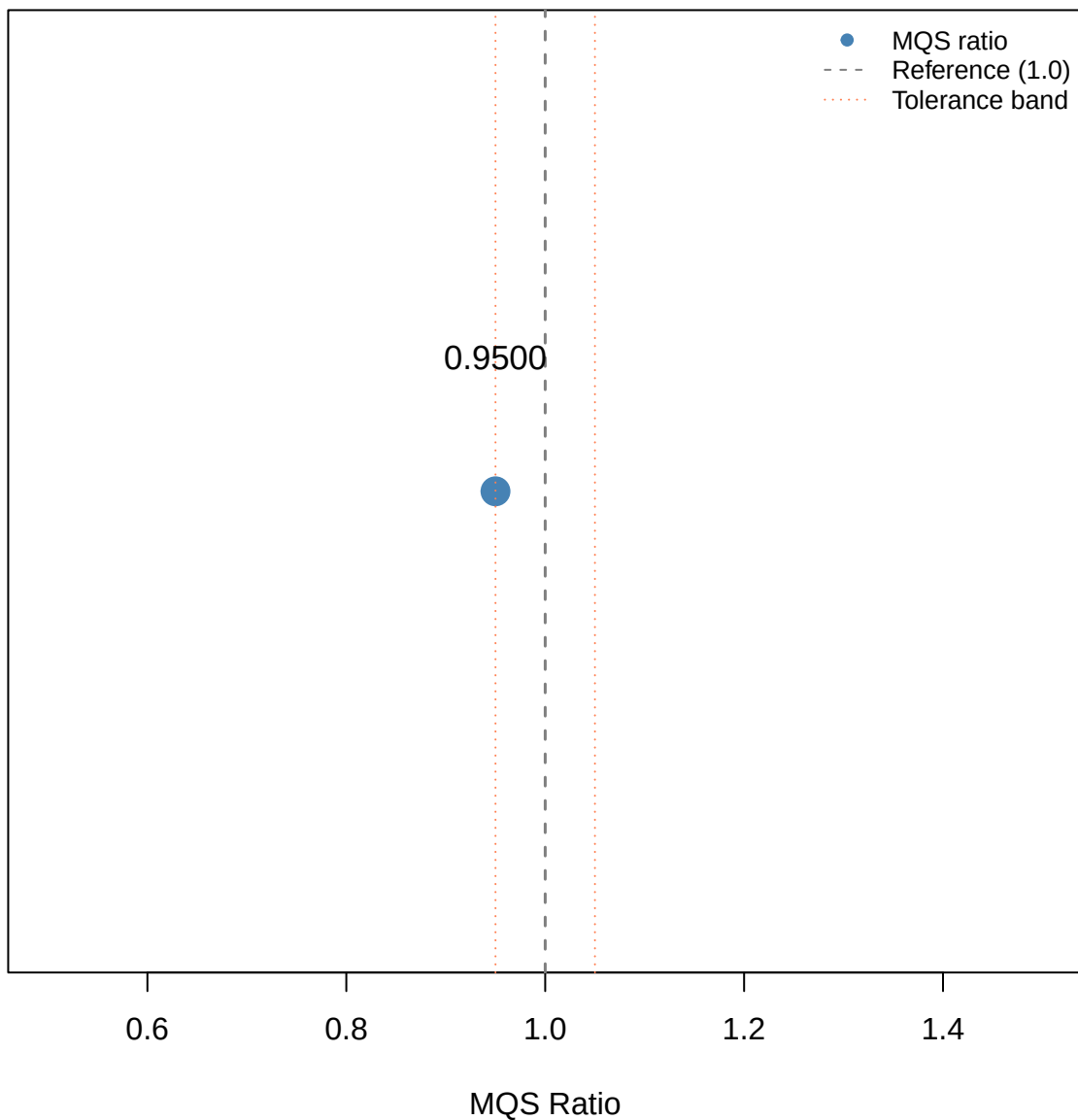
MMD Permutation Test



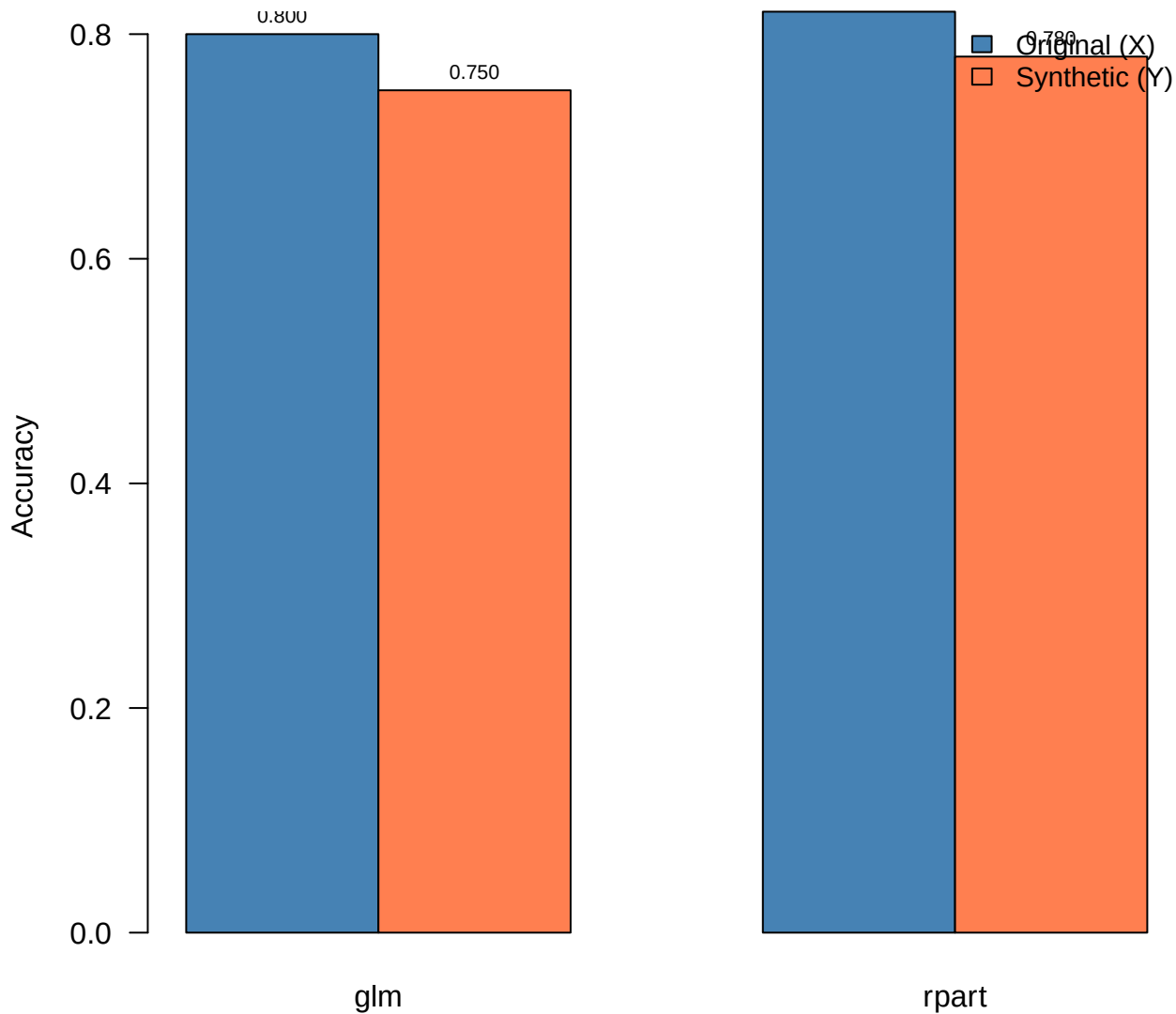
Model Performance Comparison: Accuracy



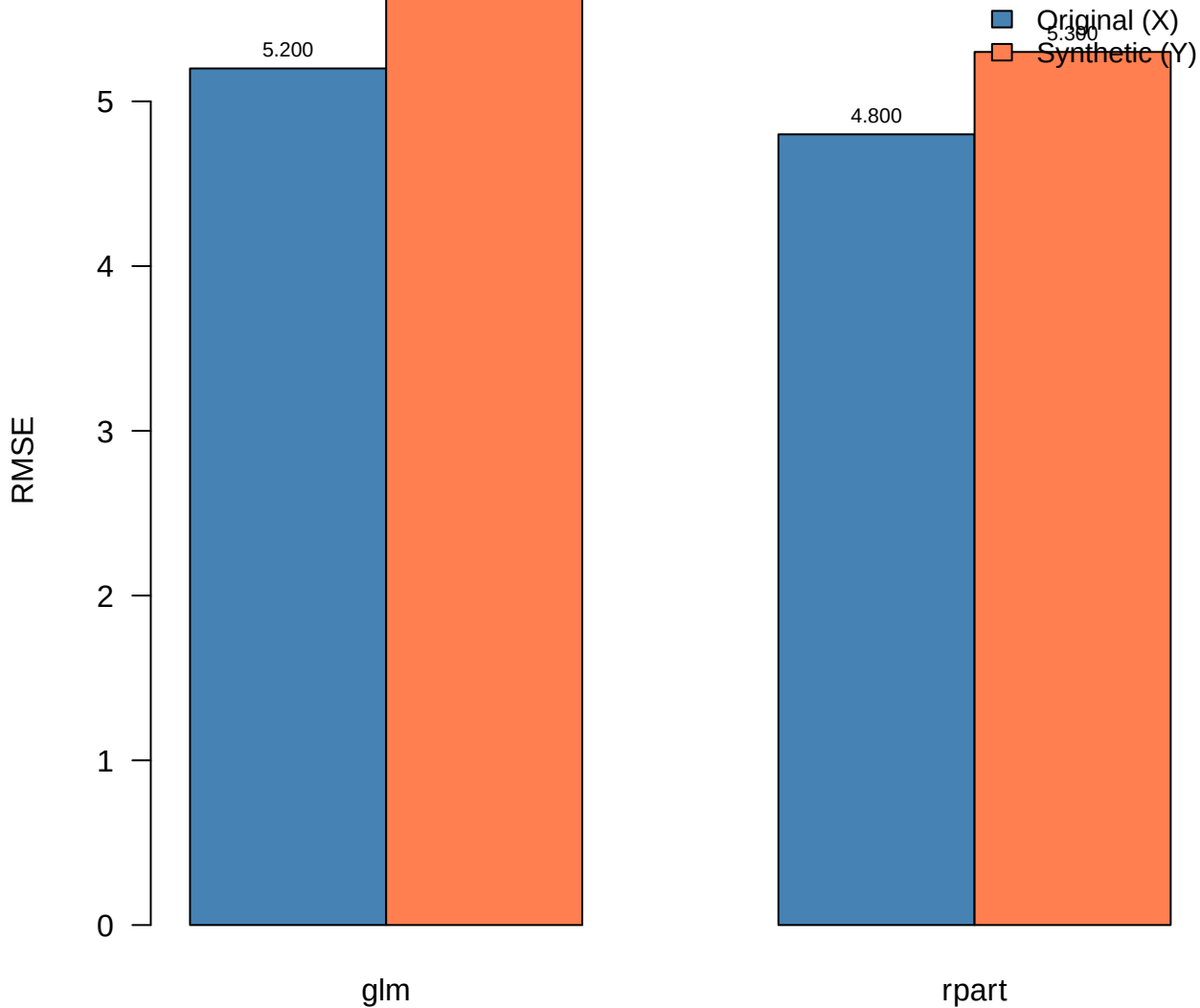
Model Quality Score Ratio



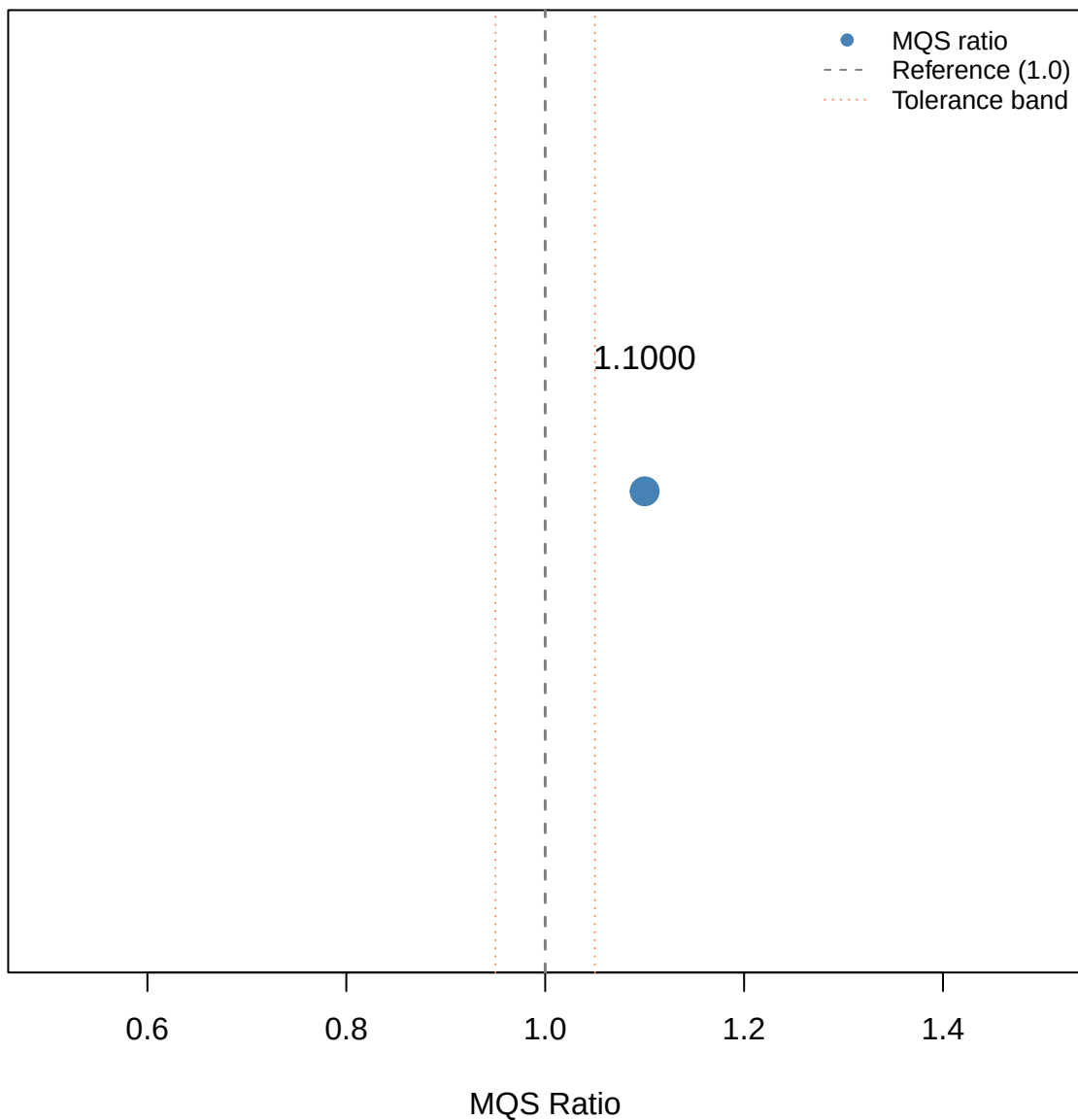
Model Performance Comparison: Accuracy



Model Performance Comparison: RMSE

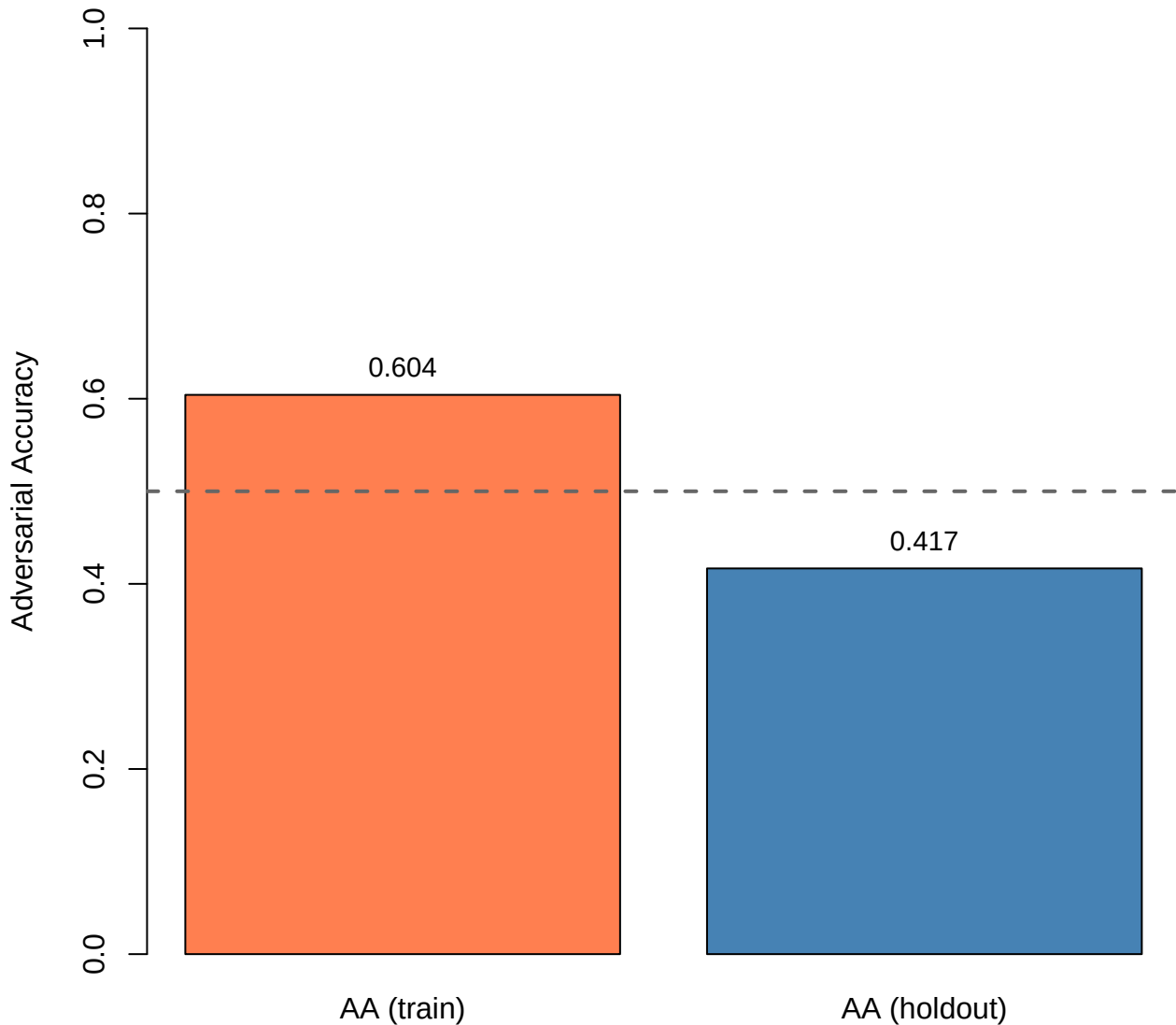


Model Quality Score Ratio

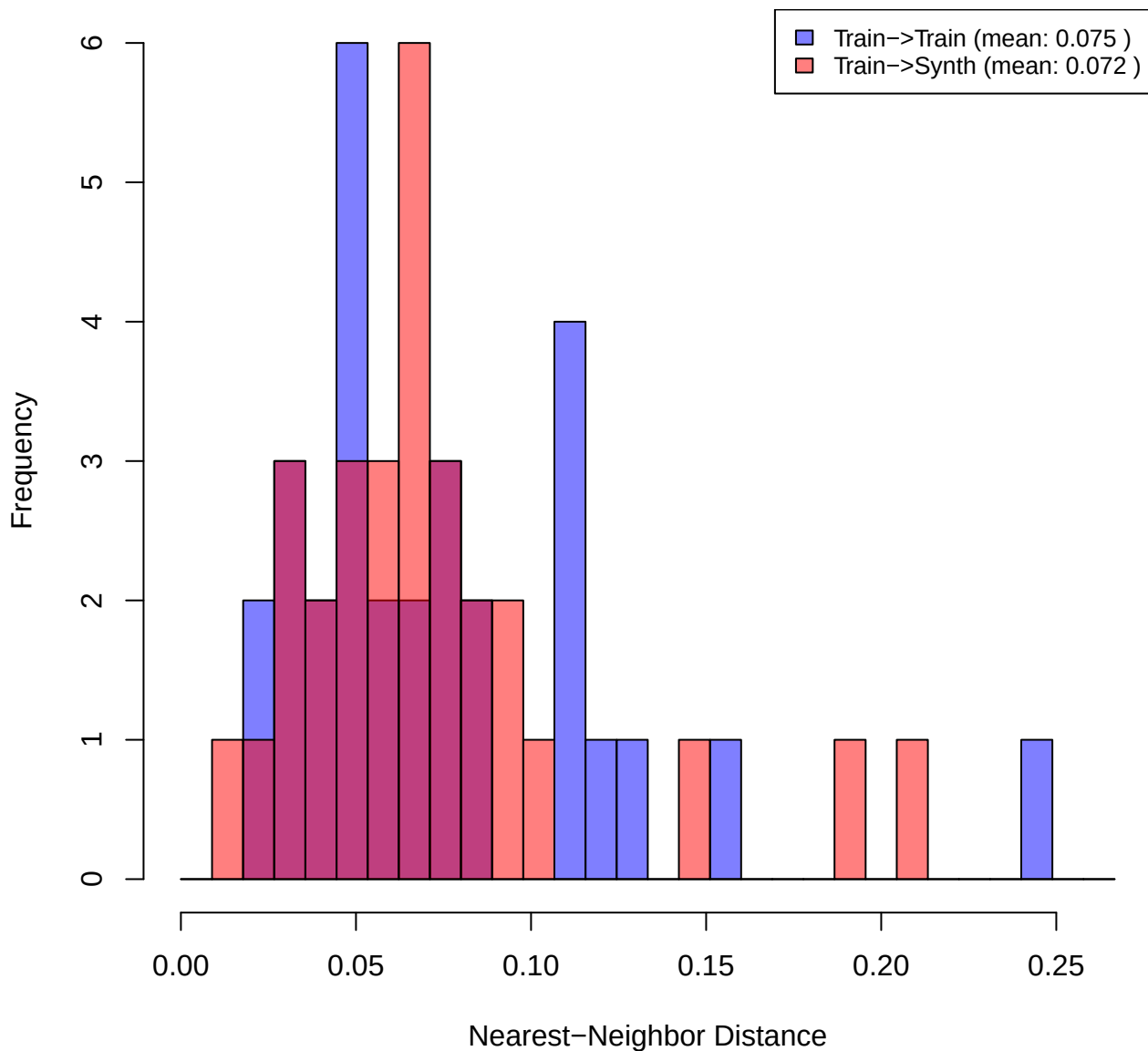


Adversarial Accuracy

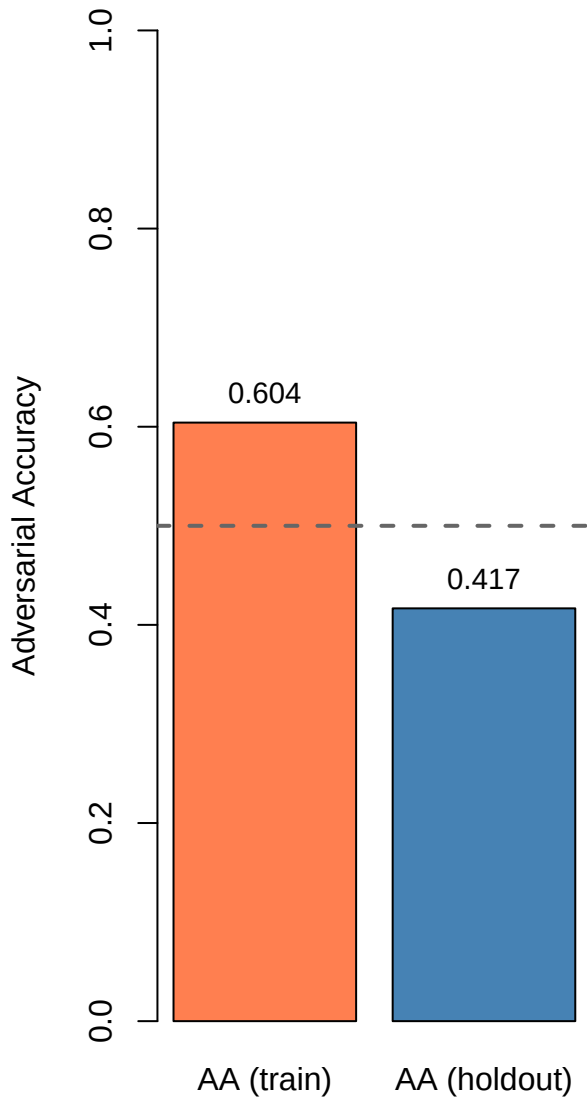
Privacy Loss: -0.1875



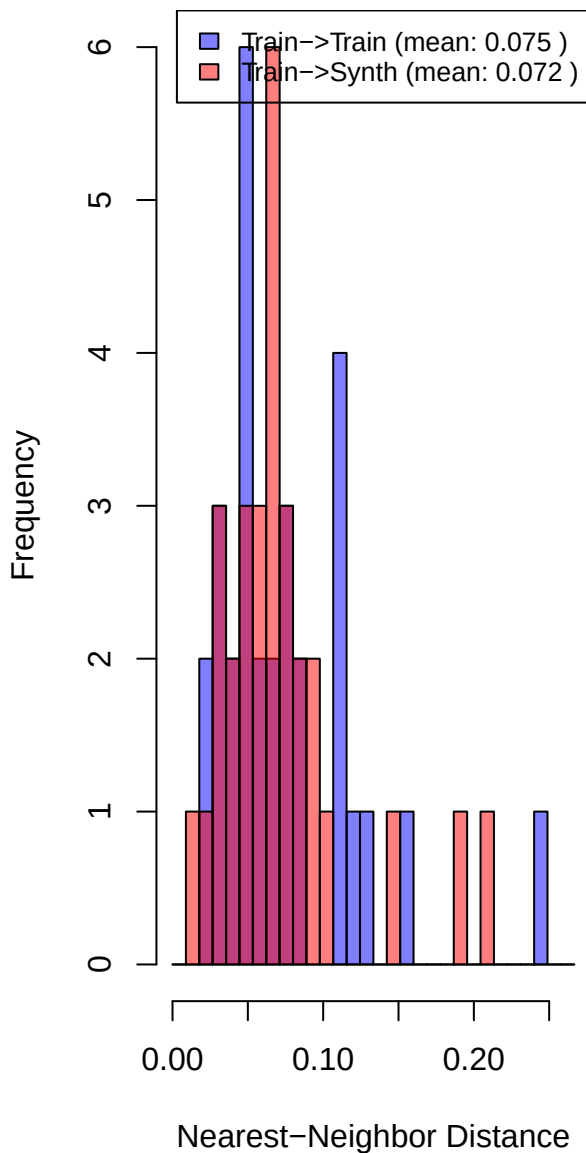
NN Distance Distributions



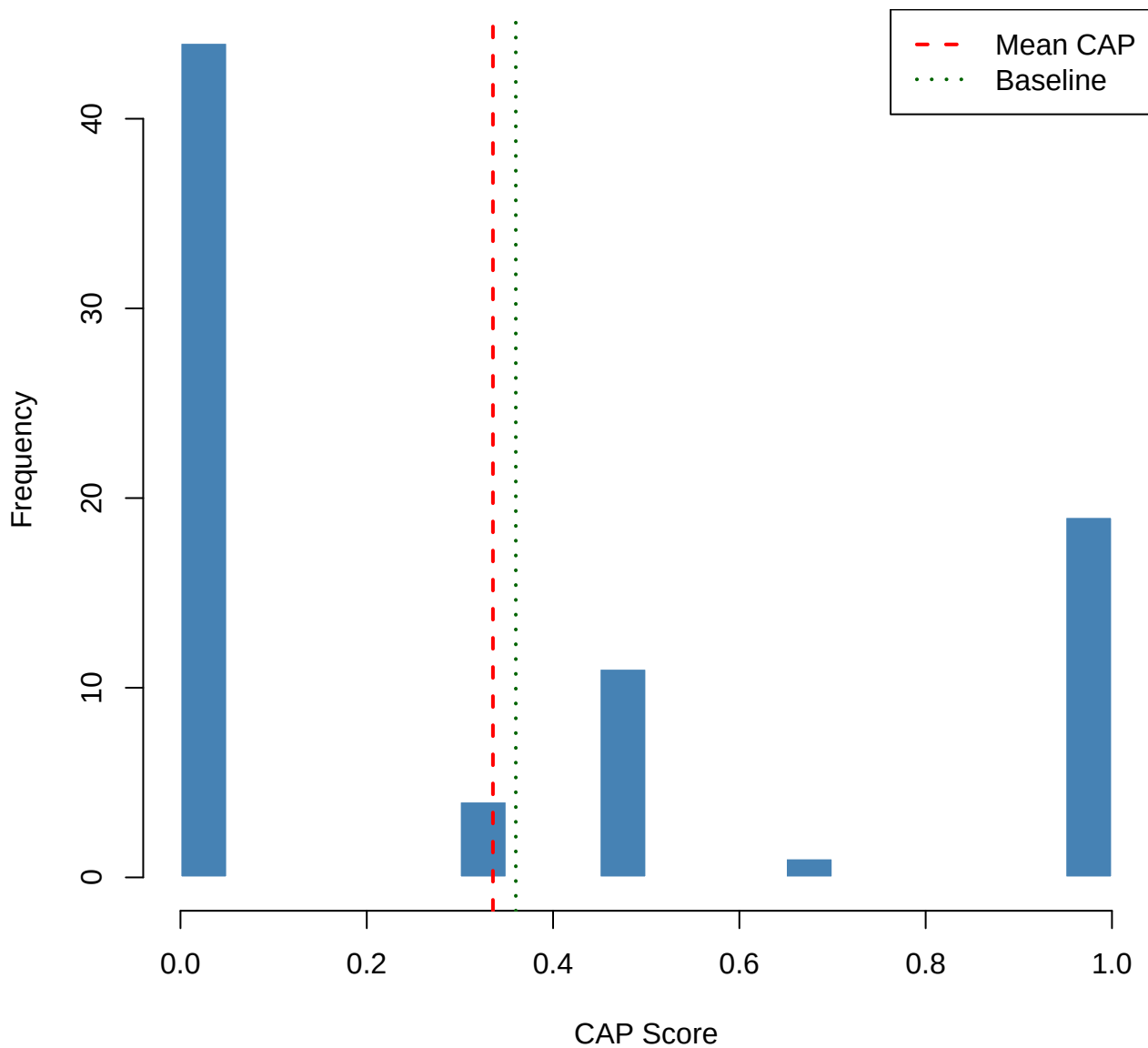
Adversarial Accuracy Privacy Loss: -0.1875



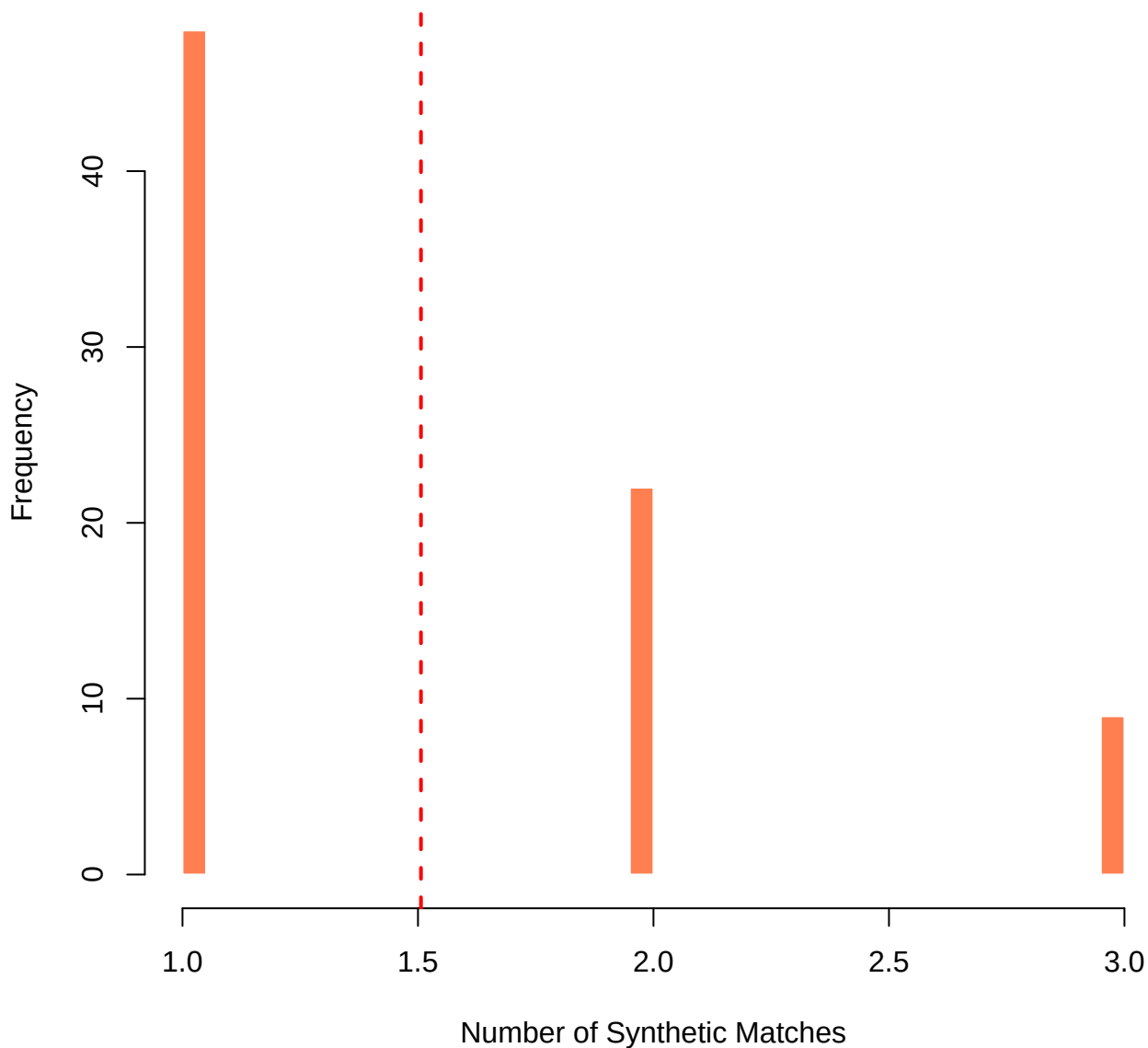
NN Distance Distributions



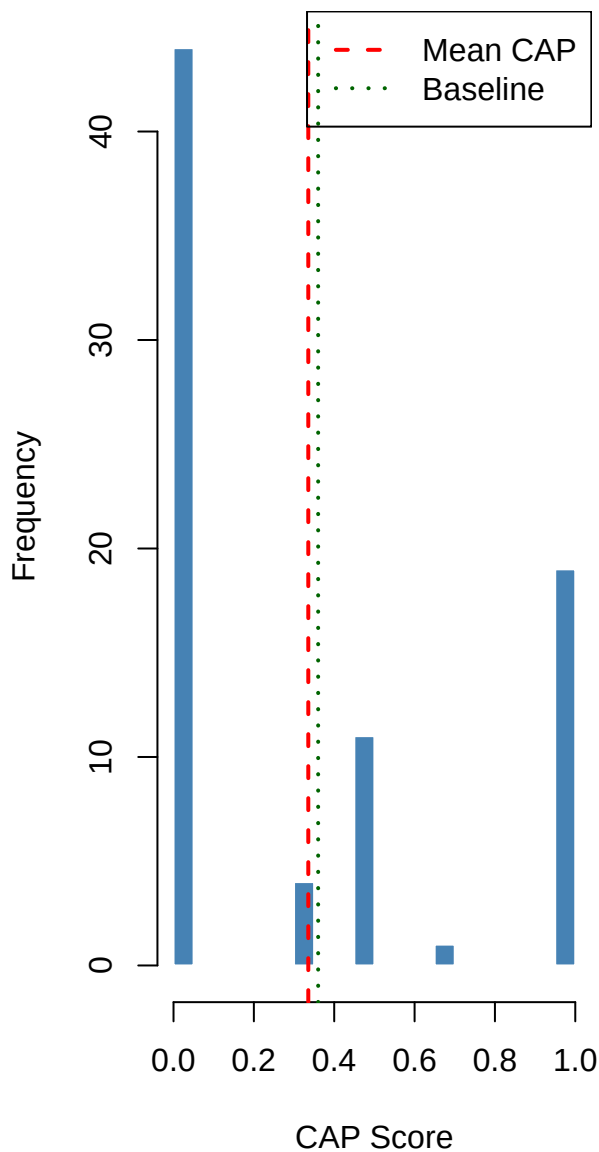
Distribution of CAP Scores



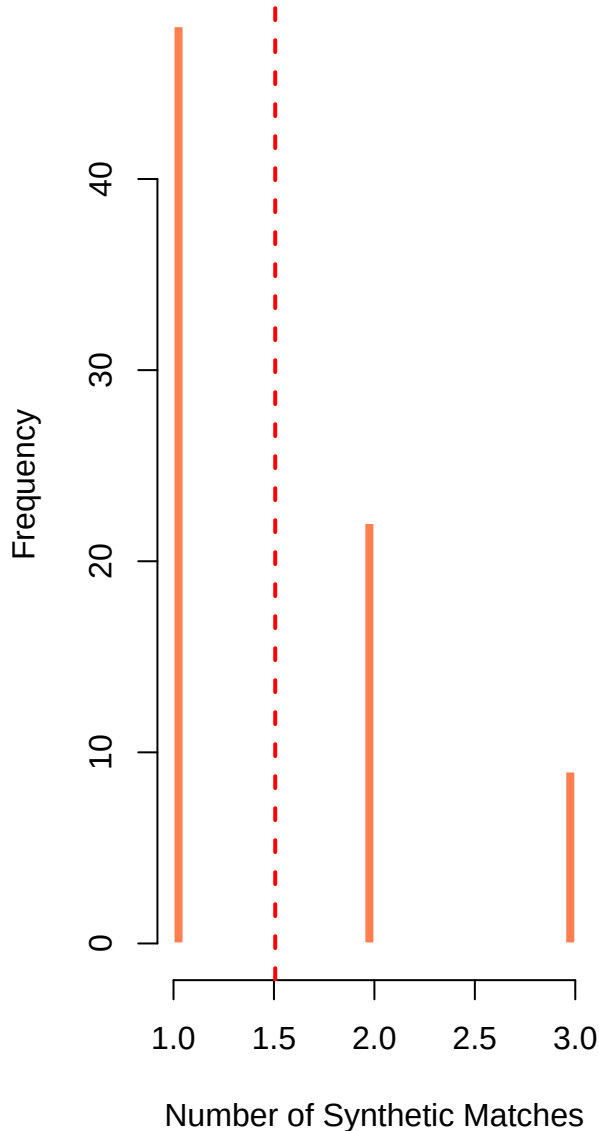
Number of Matches per Record



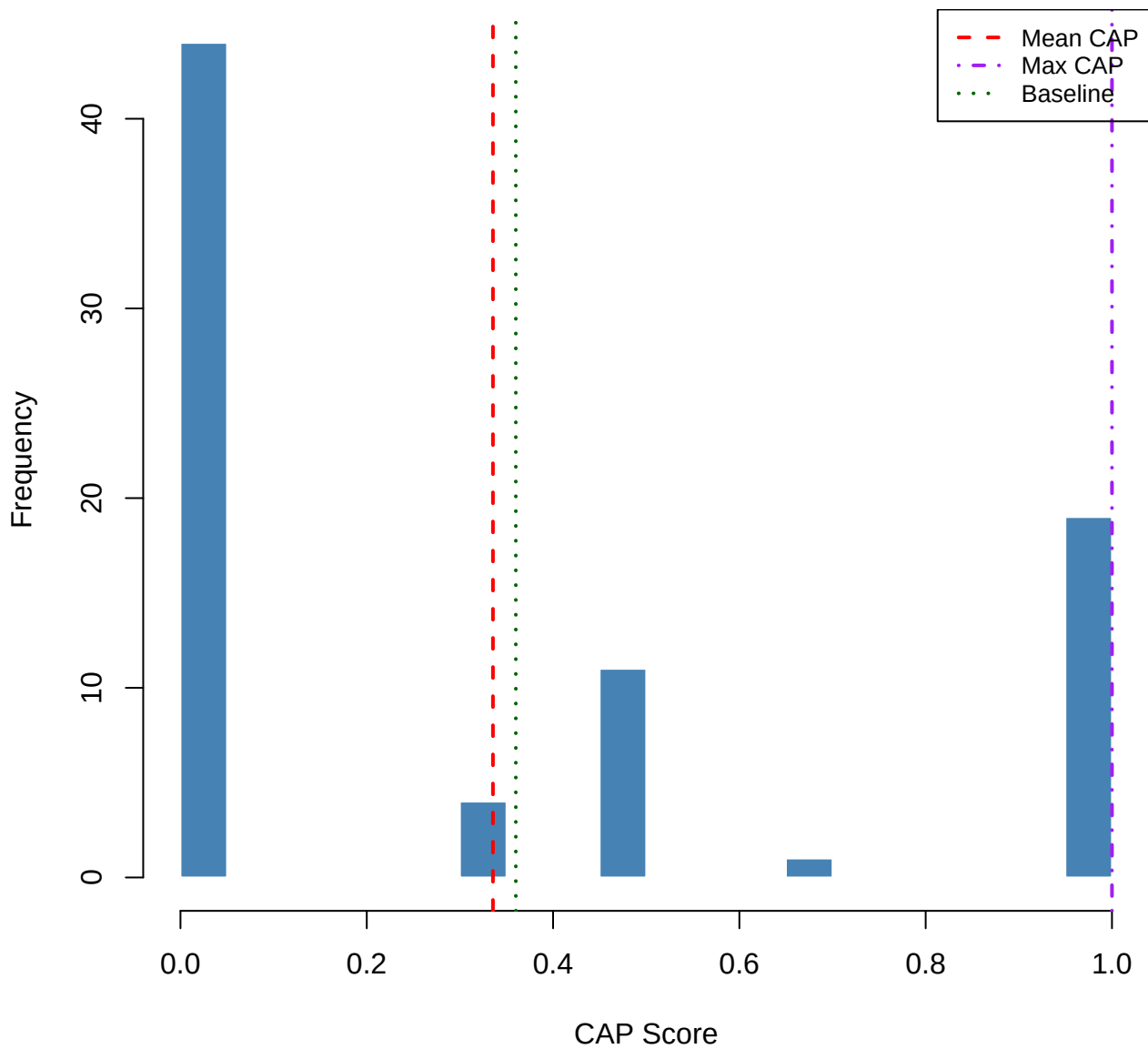
Distribution of CAP Scores



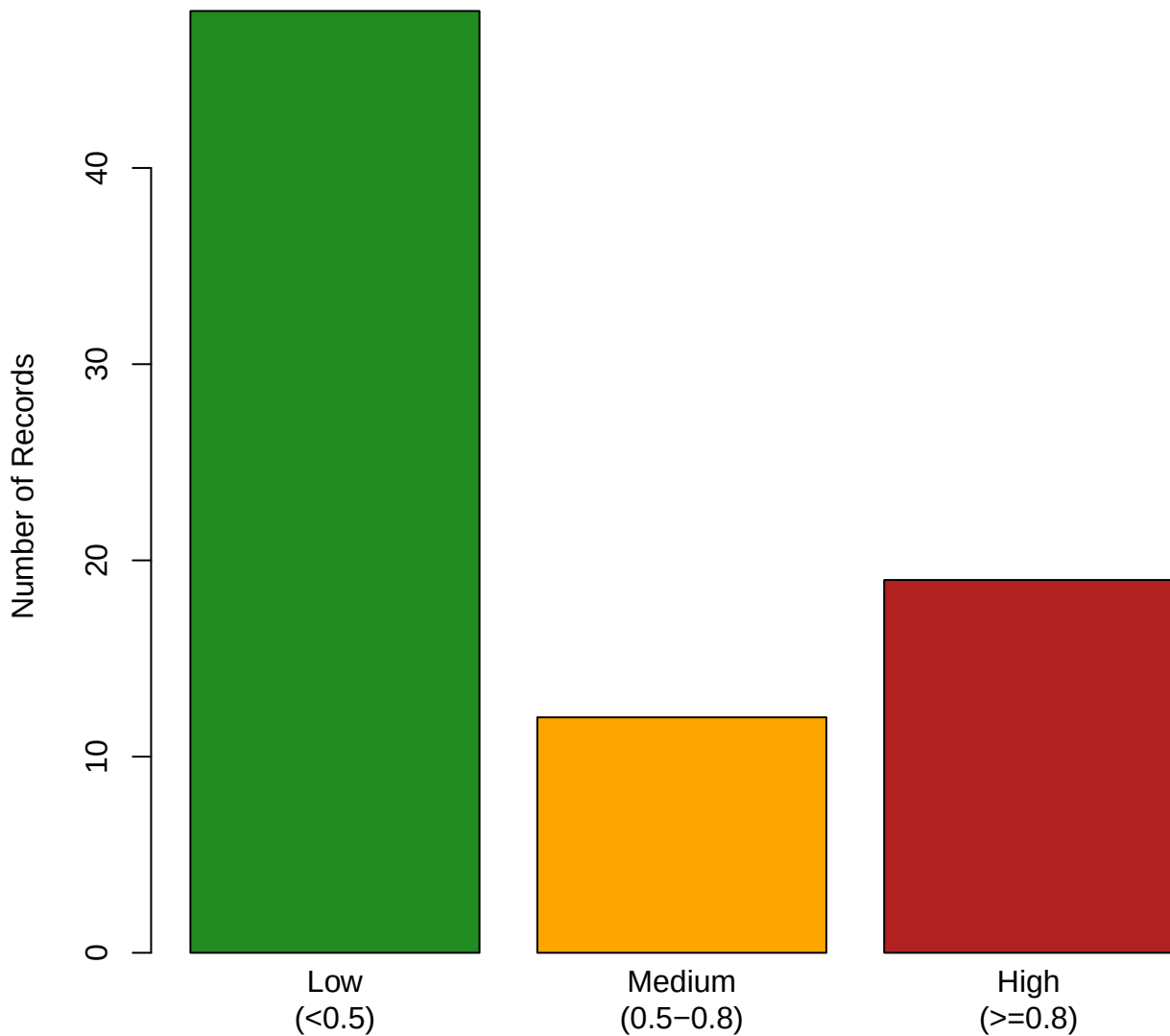
Number of Matches per Record



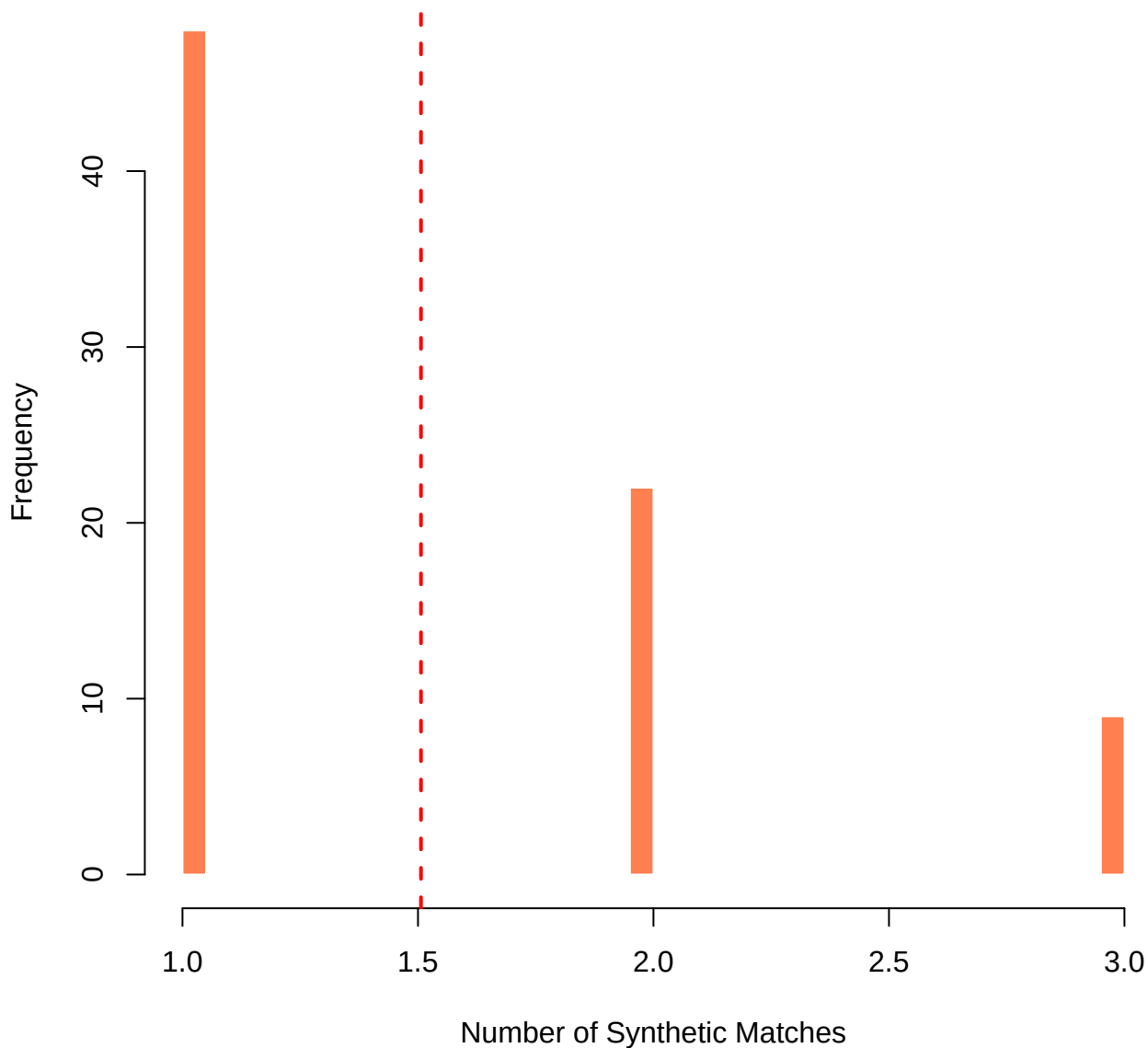
Distribution of CAP Scores



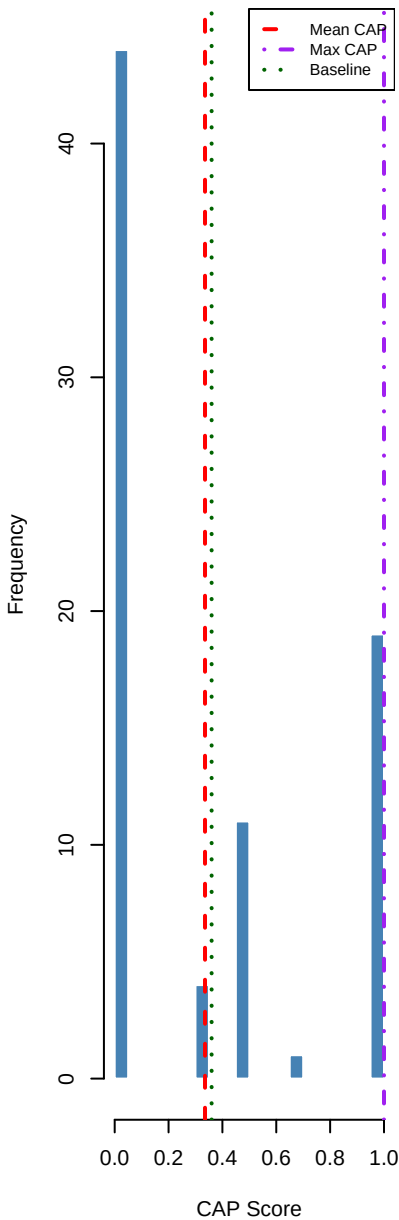
Risk Categories



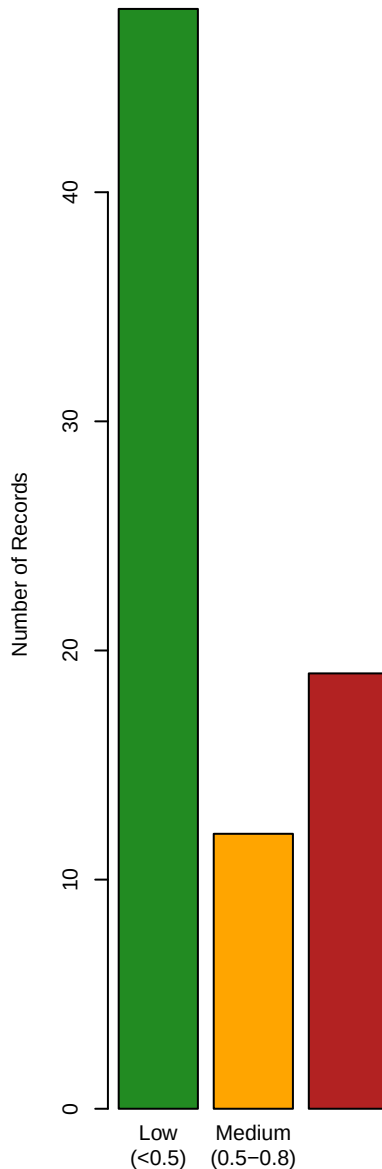
Matches per Record



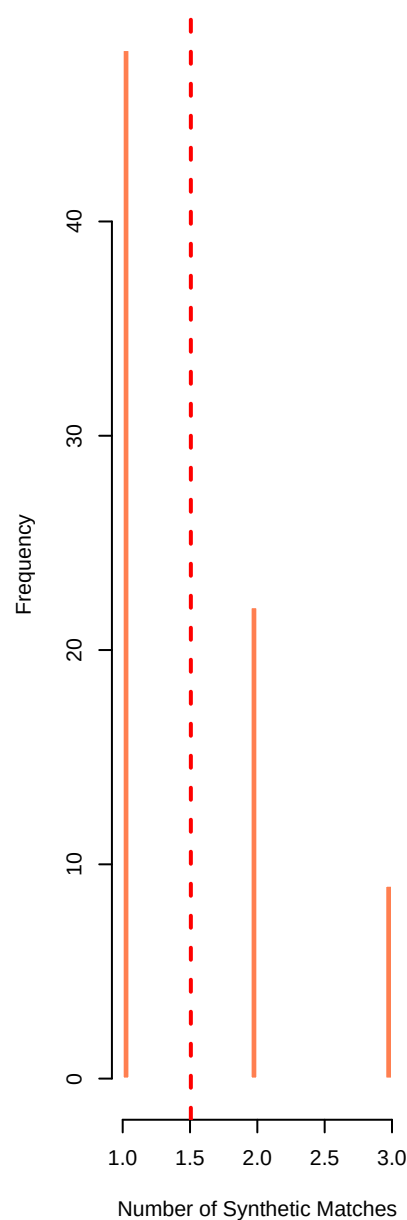
Distribution of CAP Scores



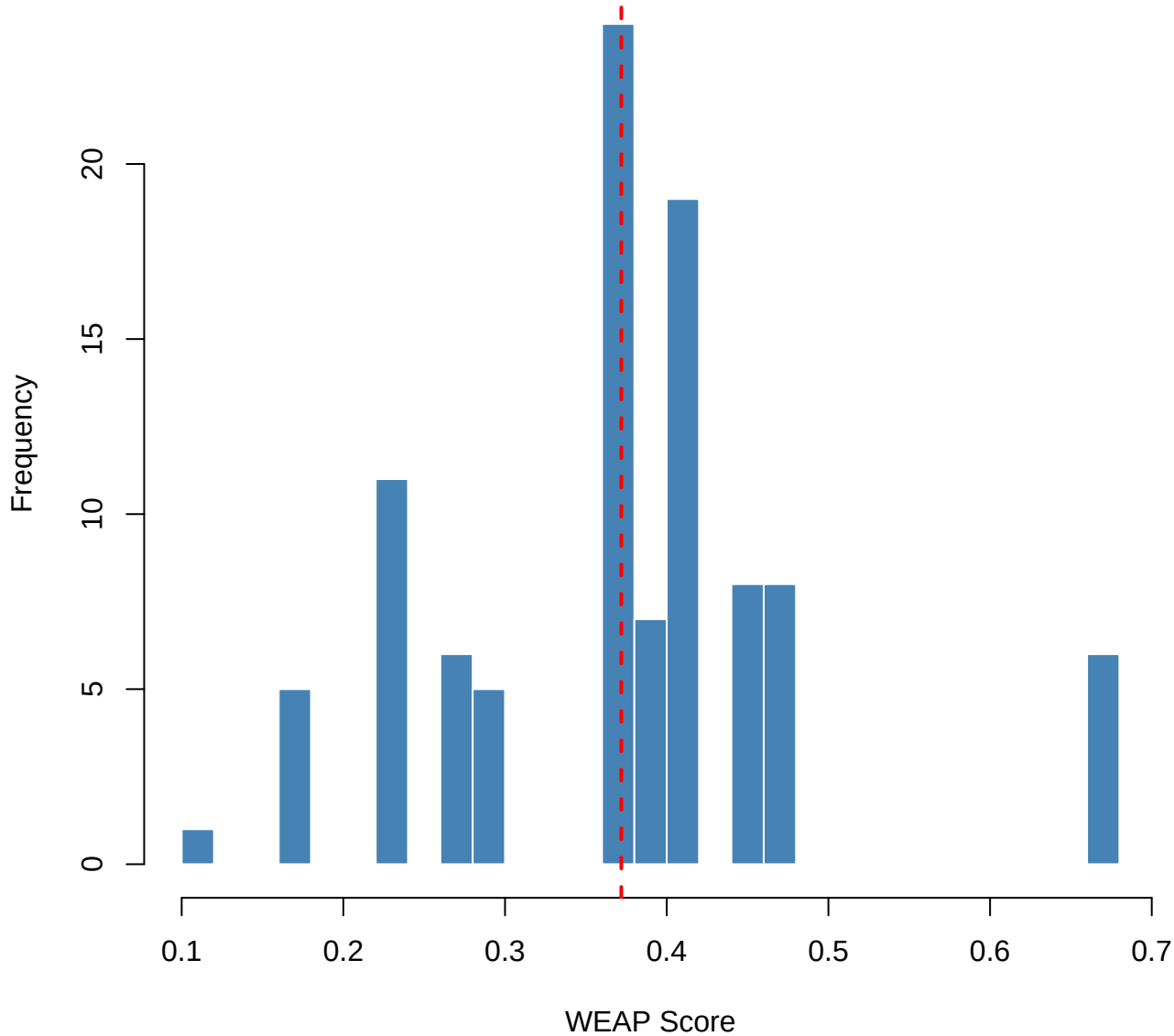
Risk Categories



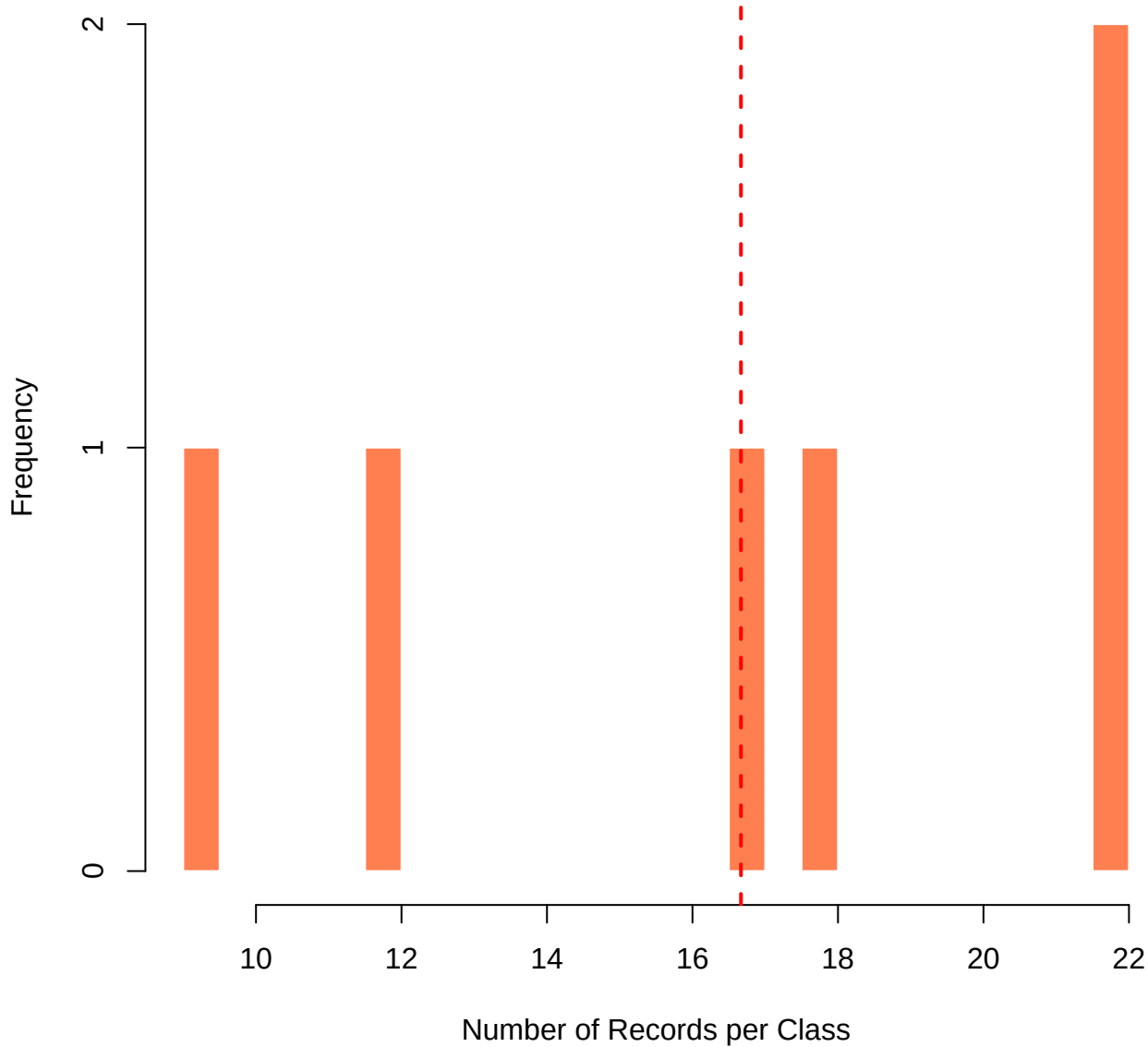
Matches per Record



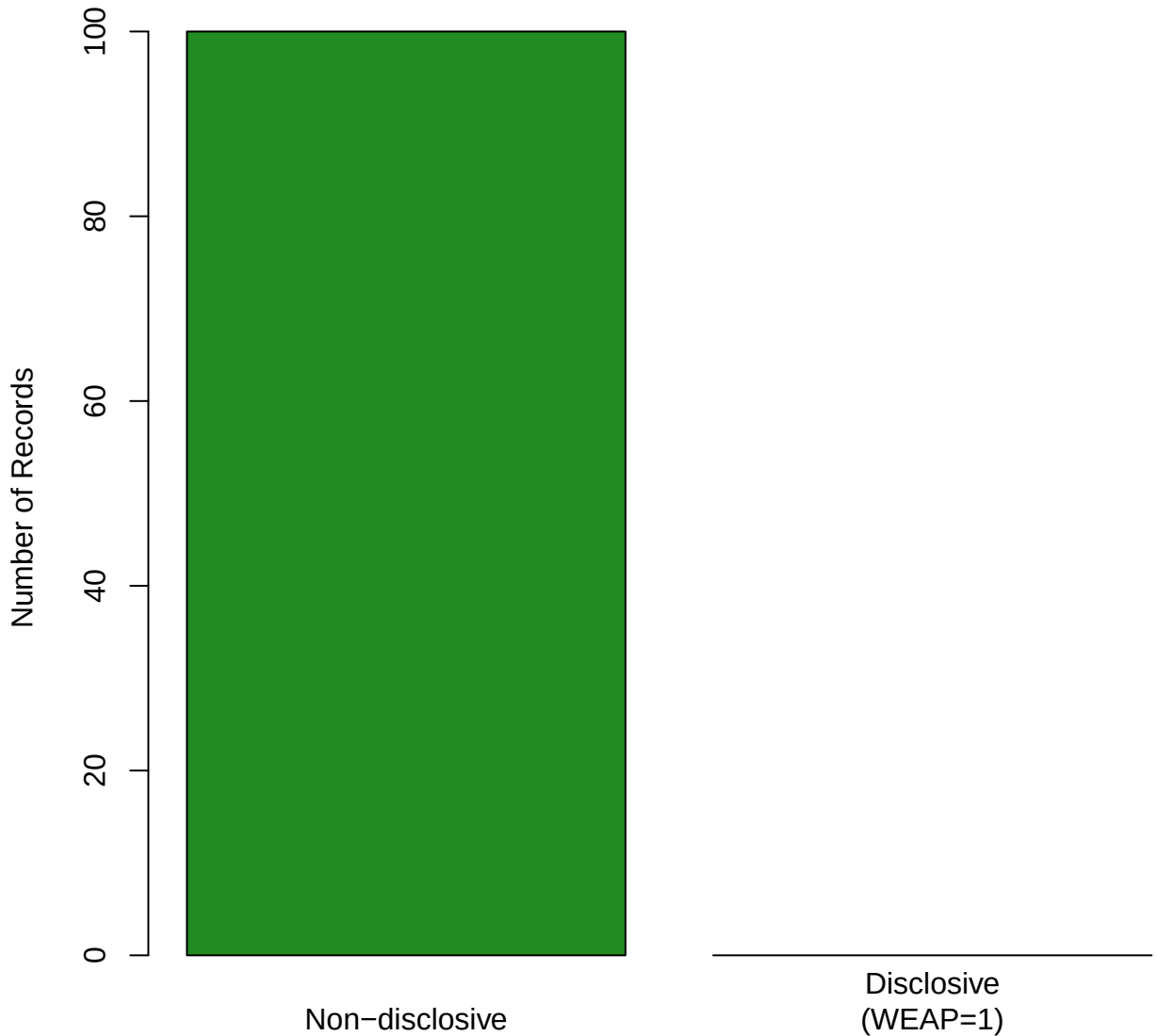
Distribution of WEAP Scores



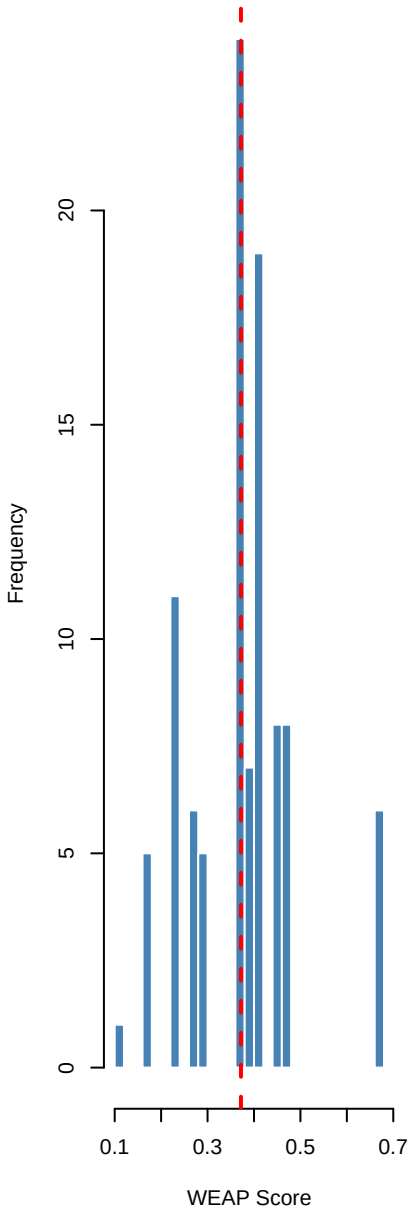
Equivalence Class Sizes



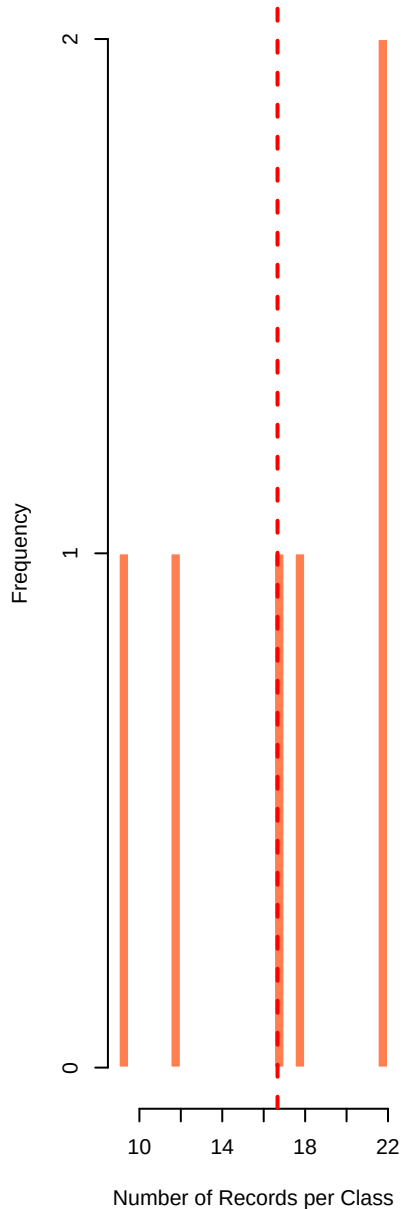
Record Disclosure Status



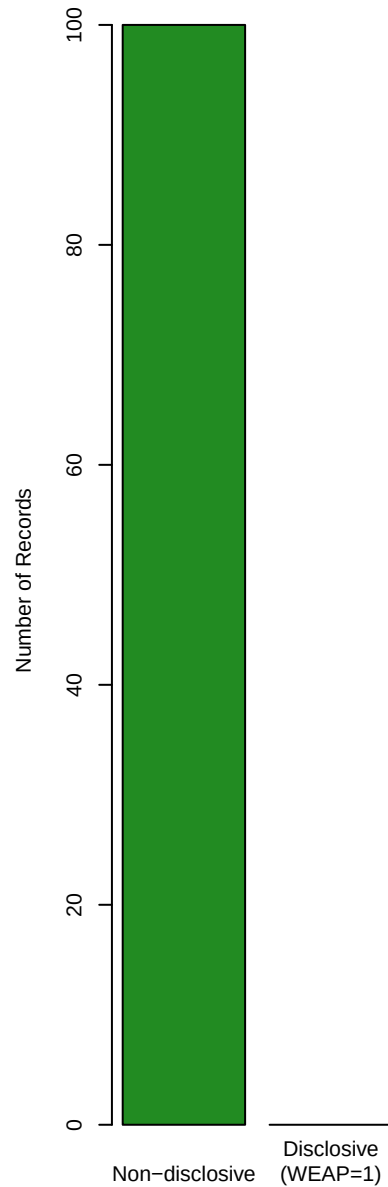
Distribution of WEAP Scores



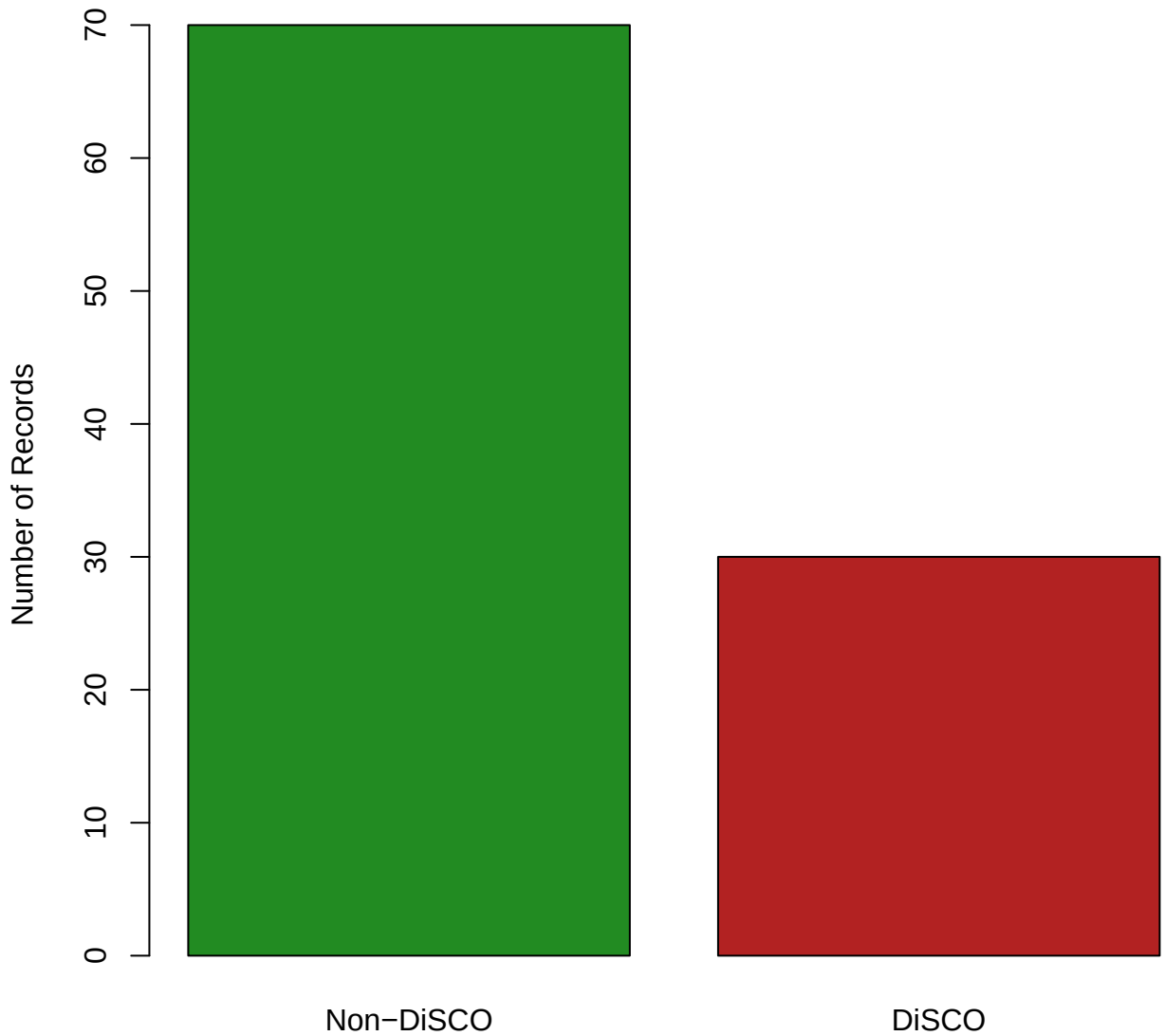
Equivalence Class Sizes



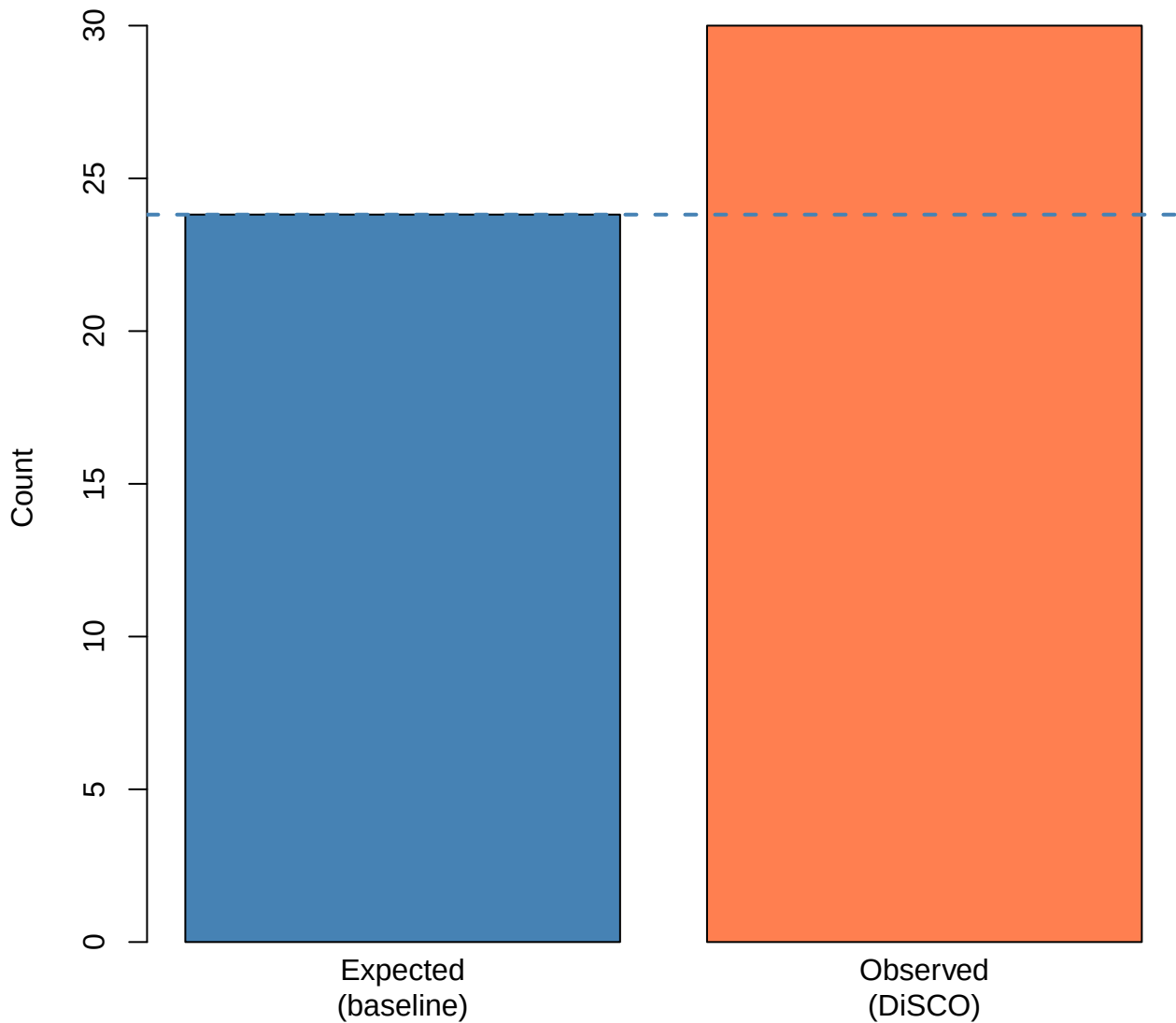
Record Disclosure Status



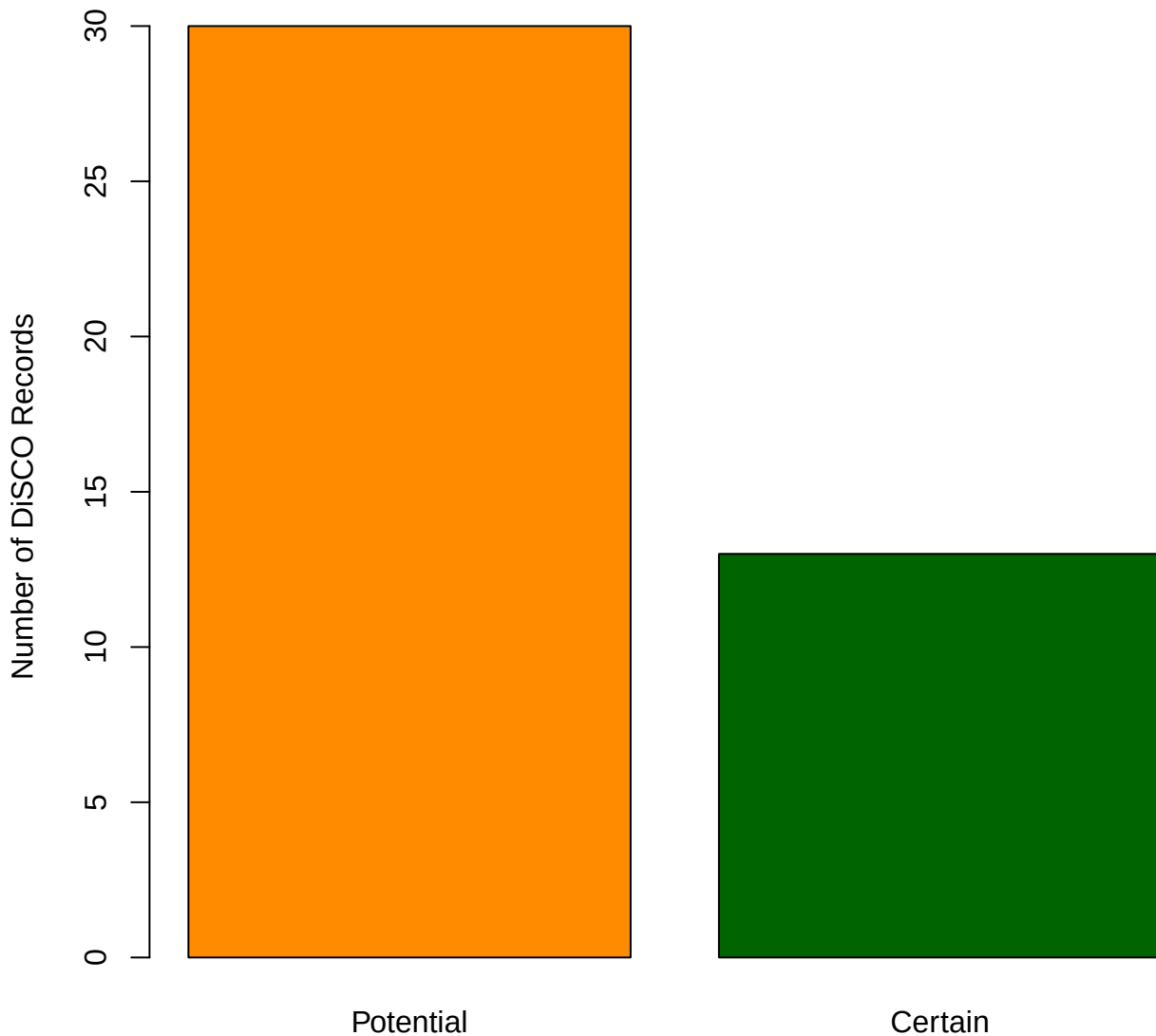
Synthetic Record Classification (potential disclosure)



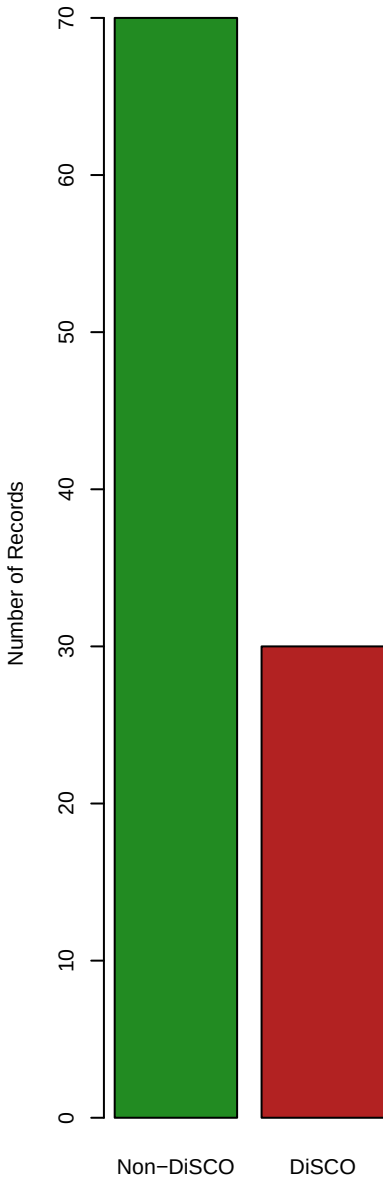
DiSCO: Observed vs Expected



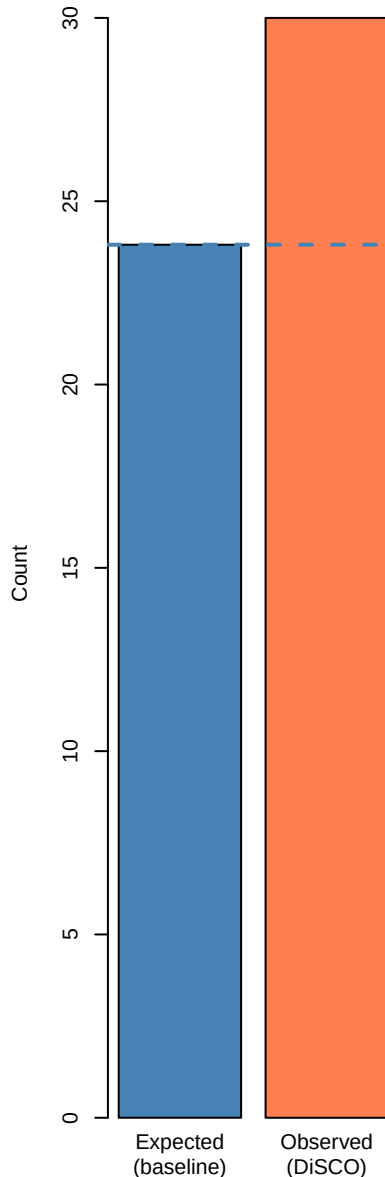
DiSCO: Potential vs Certain



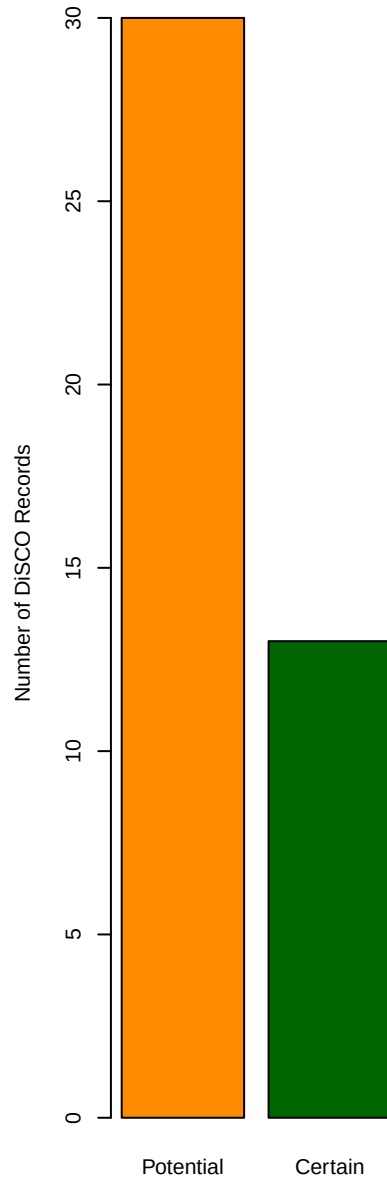
**Synthetic Record Classification
(potential disclosure)**



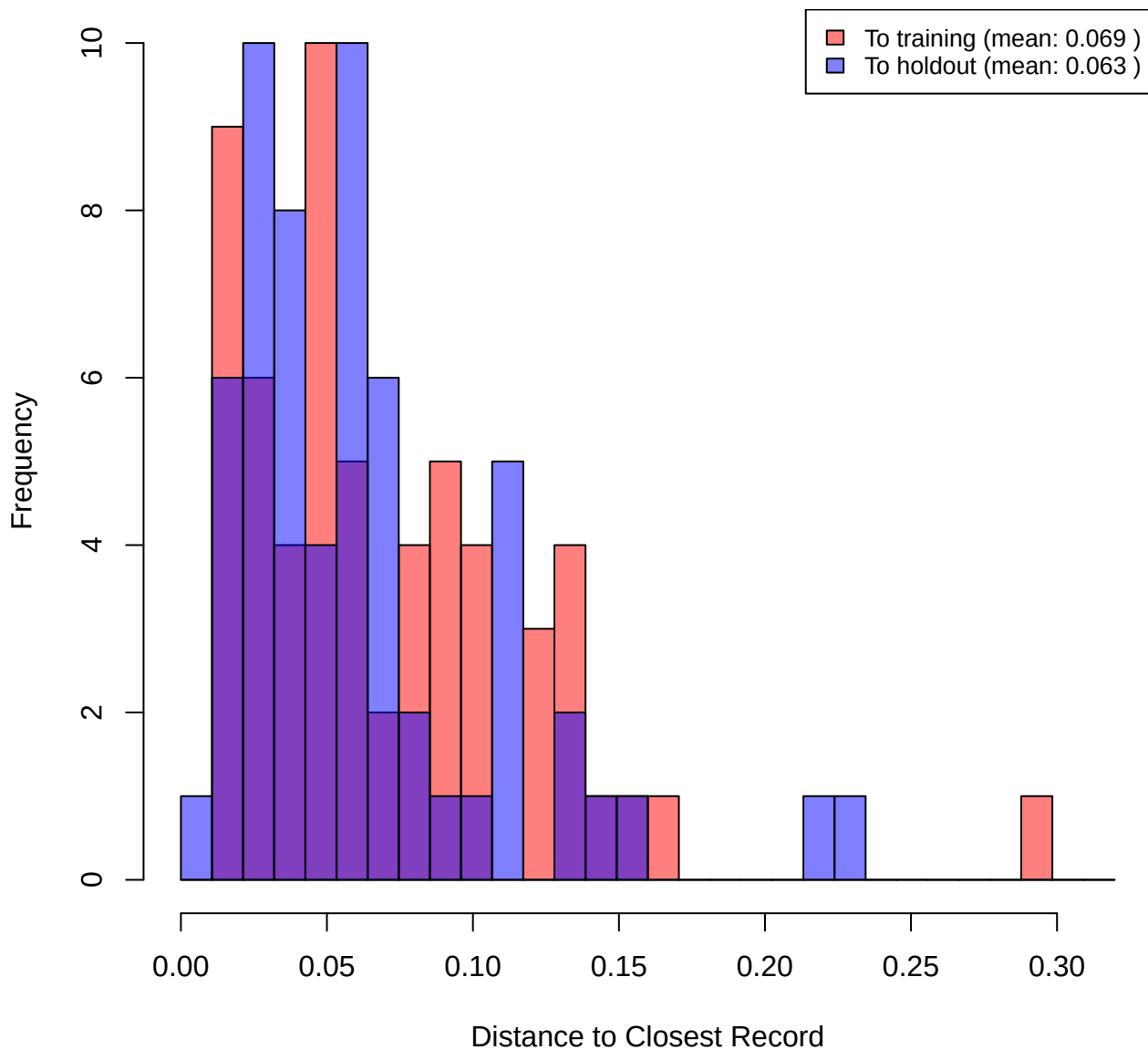
DiSCO: Observed vs Expected



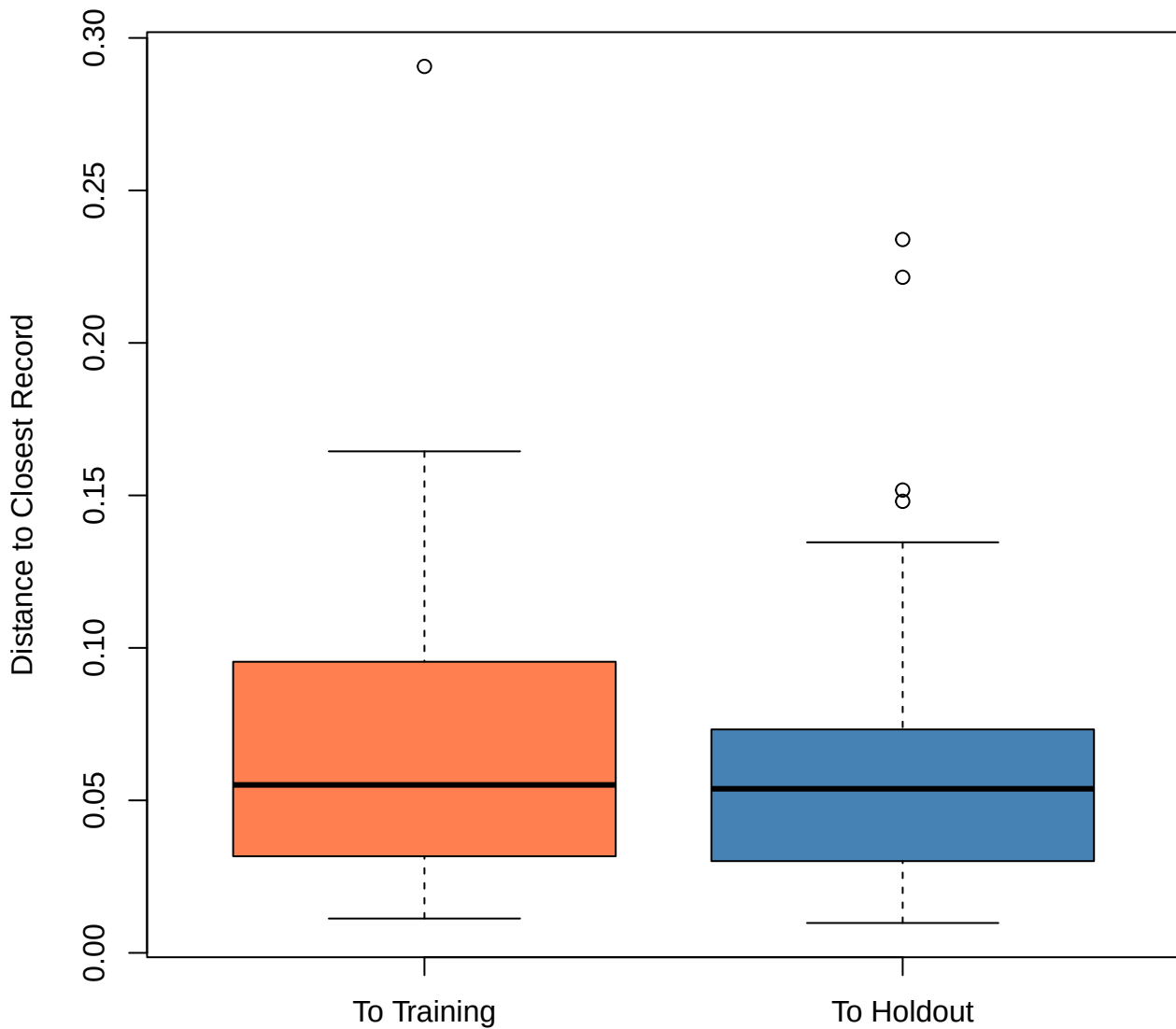
DiSCO: Potential vs Certain



DCR Distribution

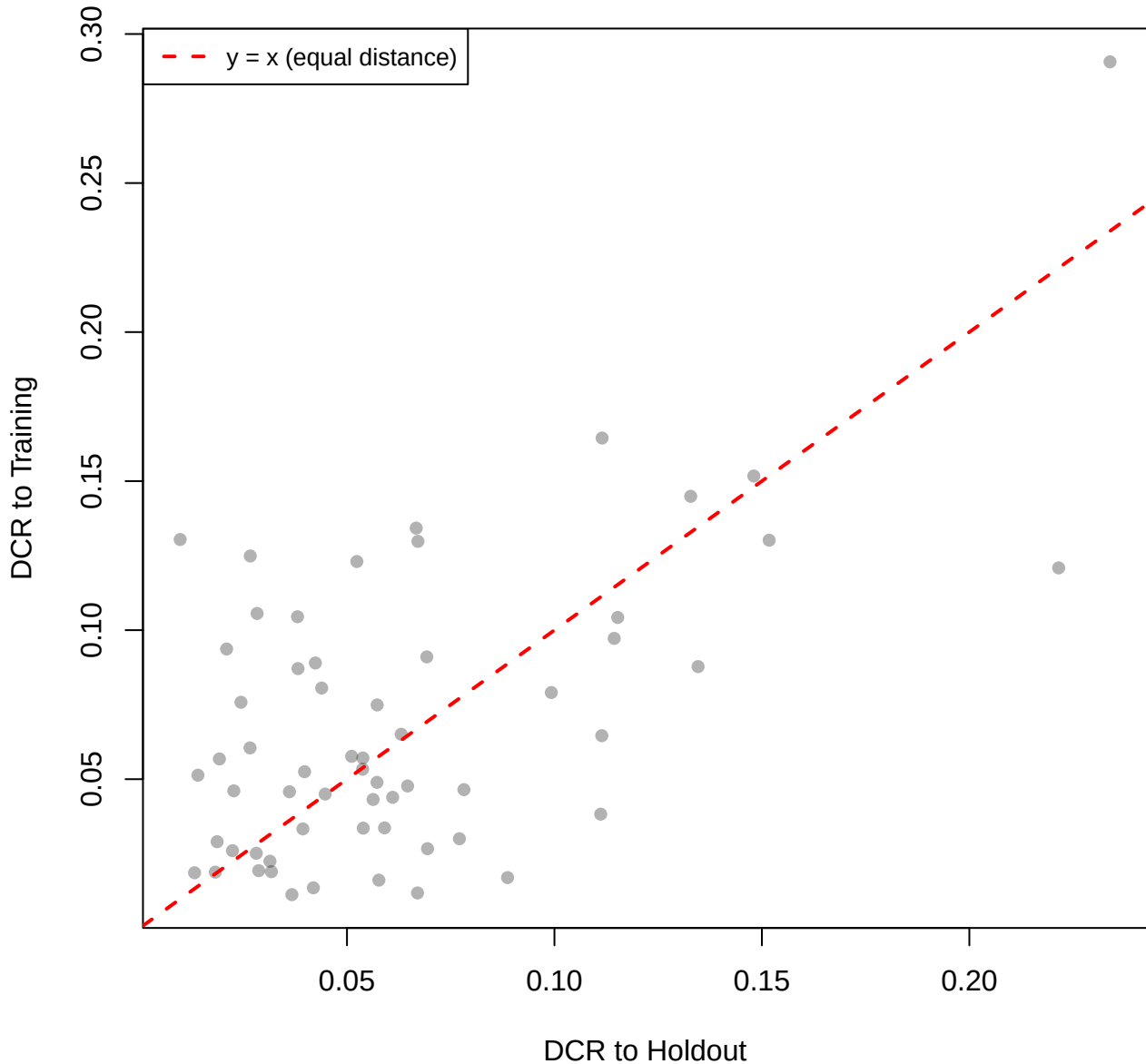


DCR Comparison

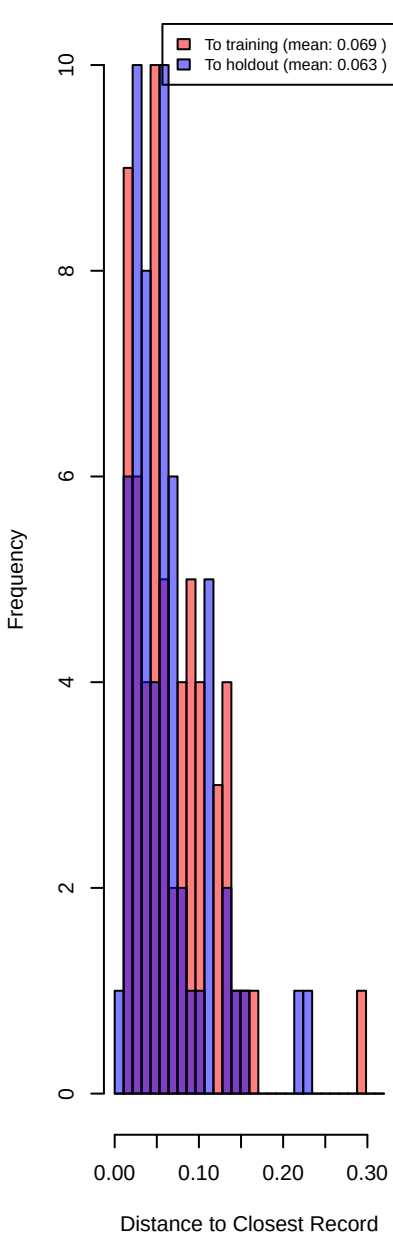


DCR: Training vs Holdout

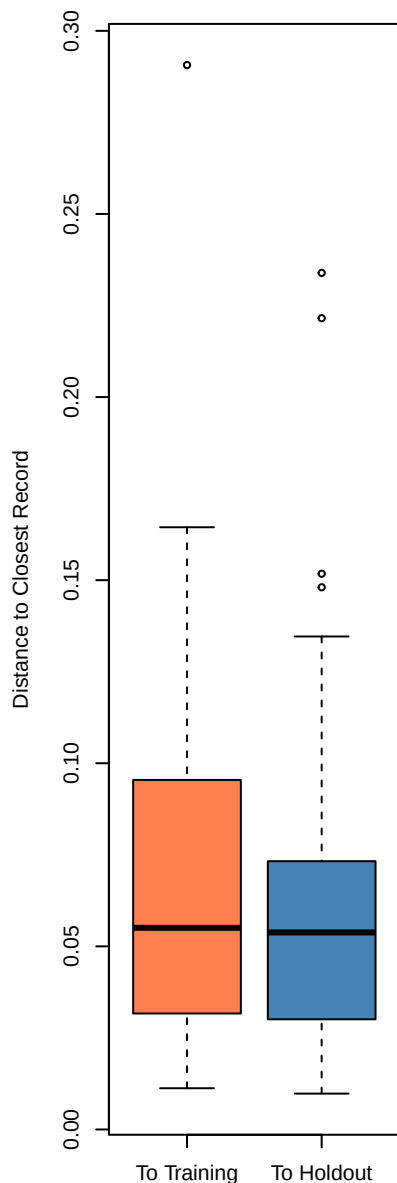
Share closer to train: 46.7%



DCR Distribution

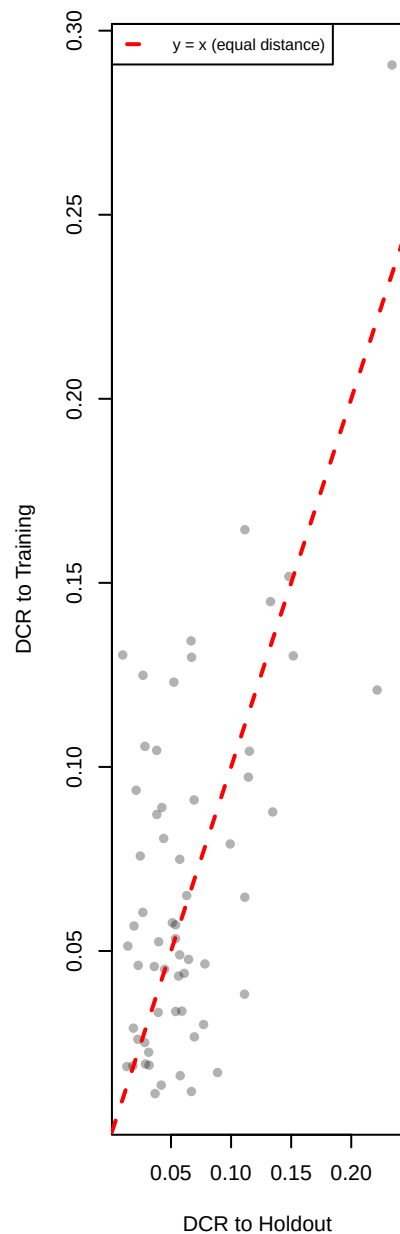


DCR Comparison

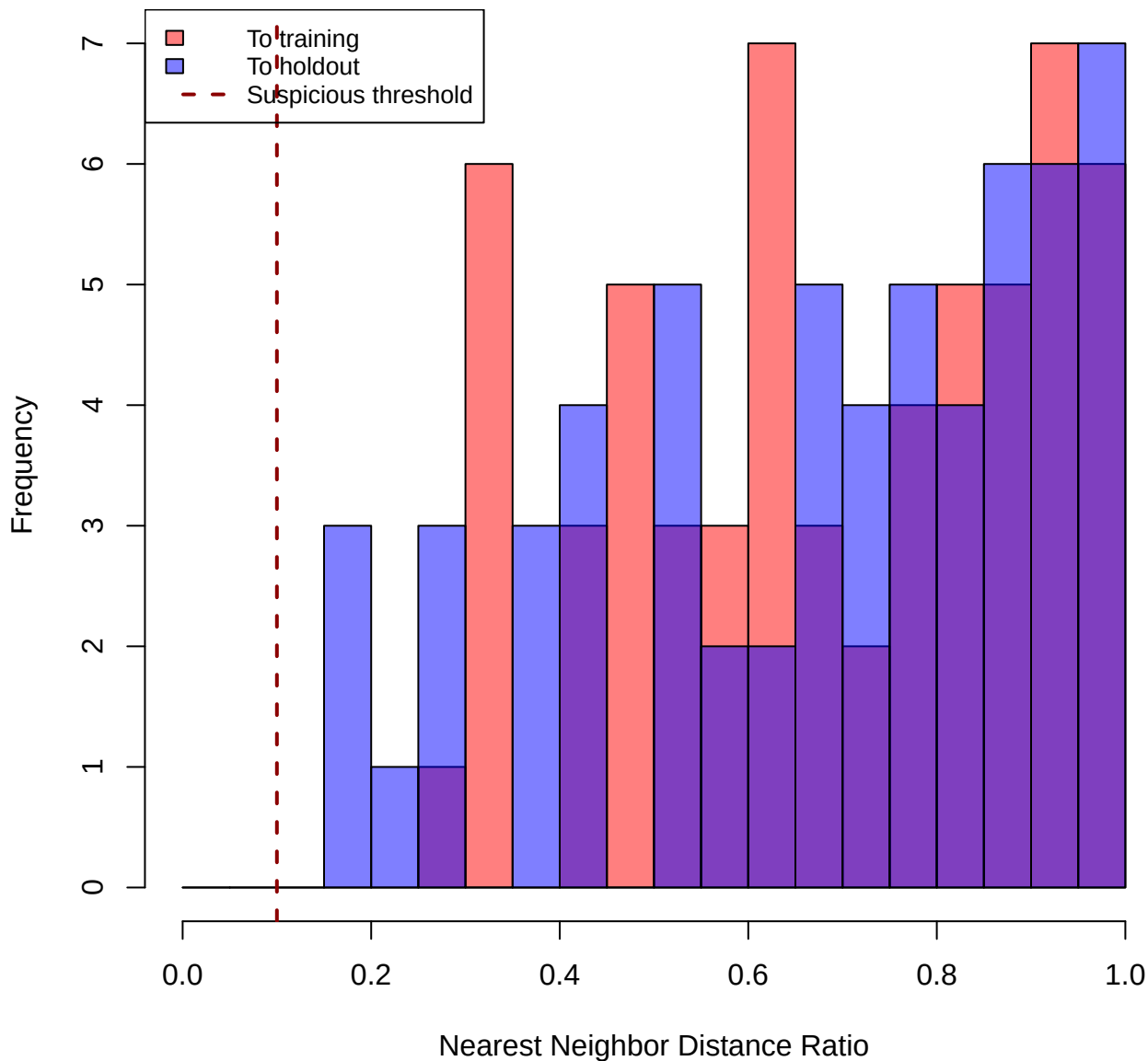


DCR: Training vs Holdout

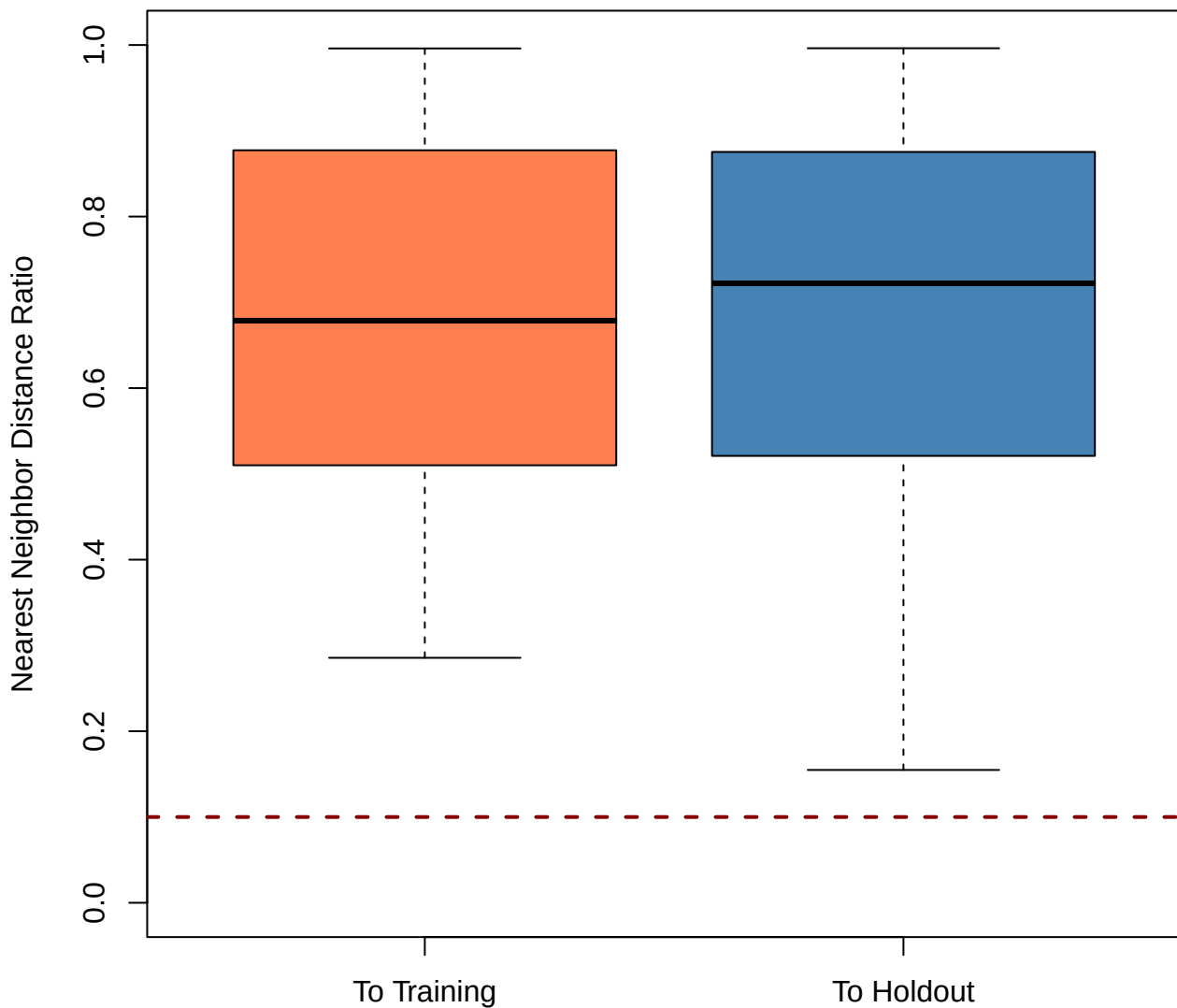
Share closer to train: 46.7%



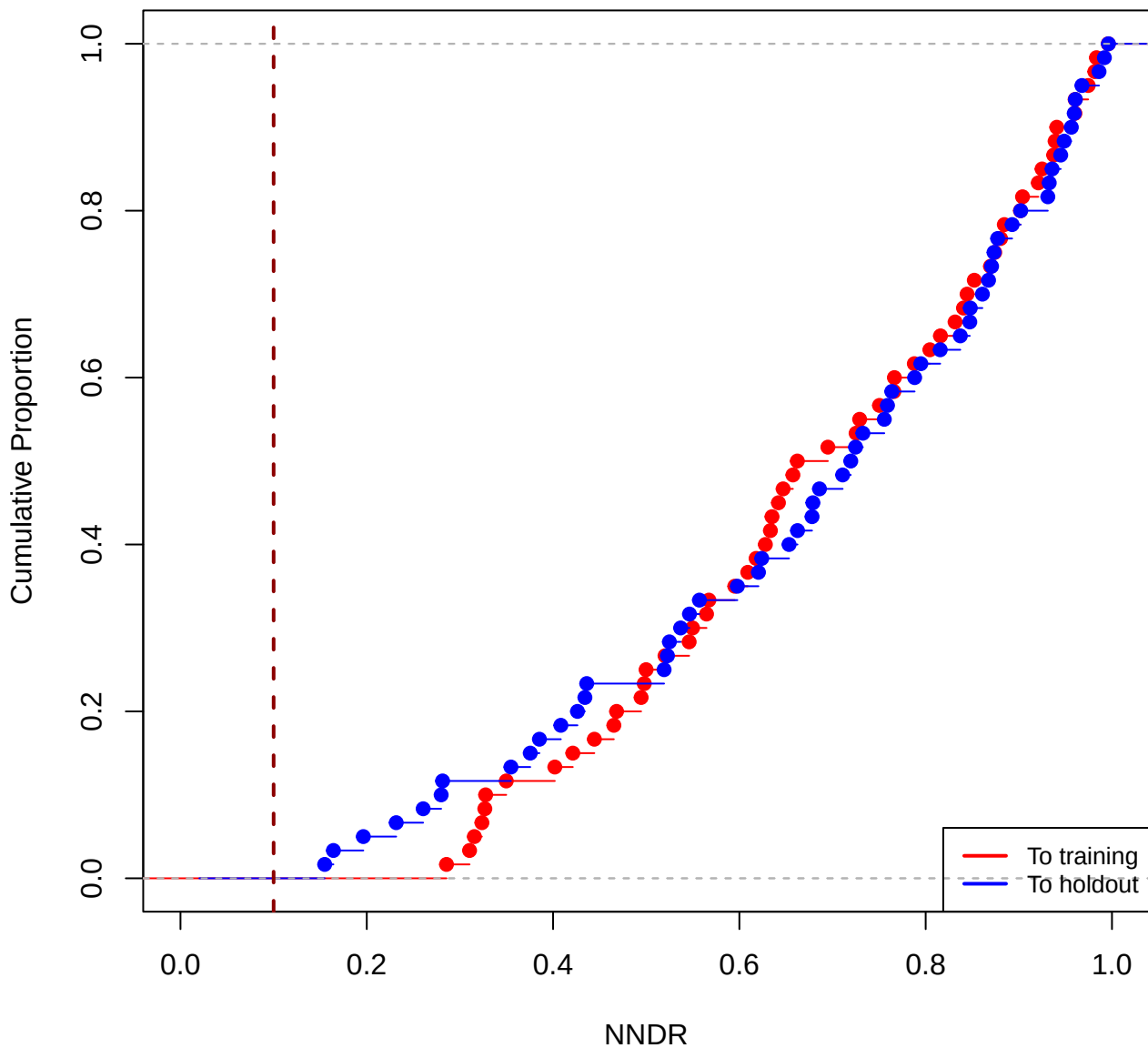
NNDR Distribution



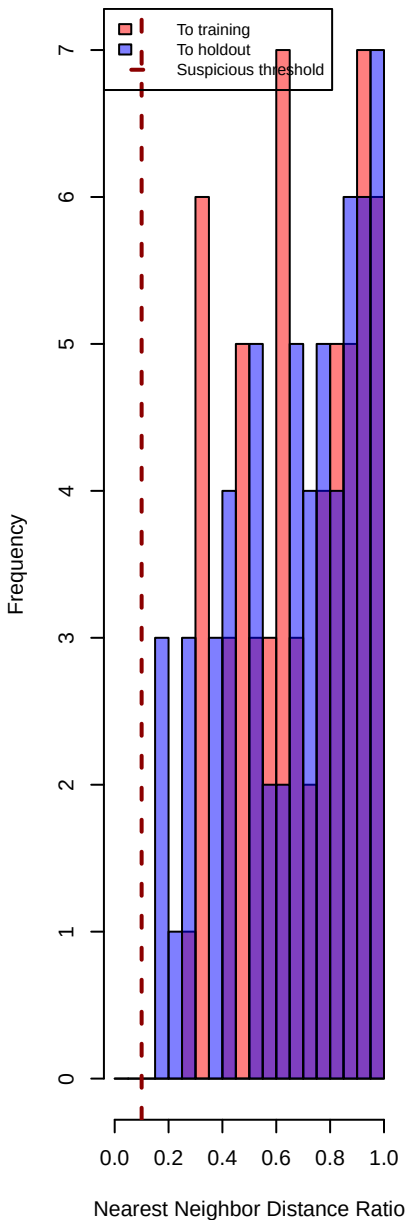
NNDR Comparison



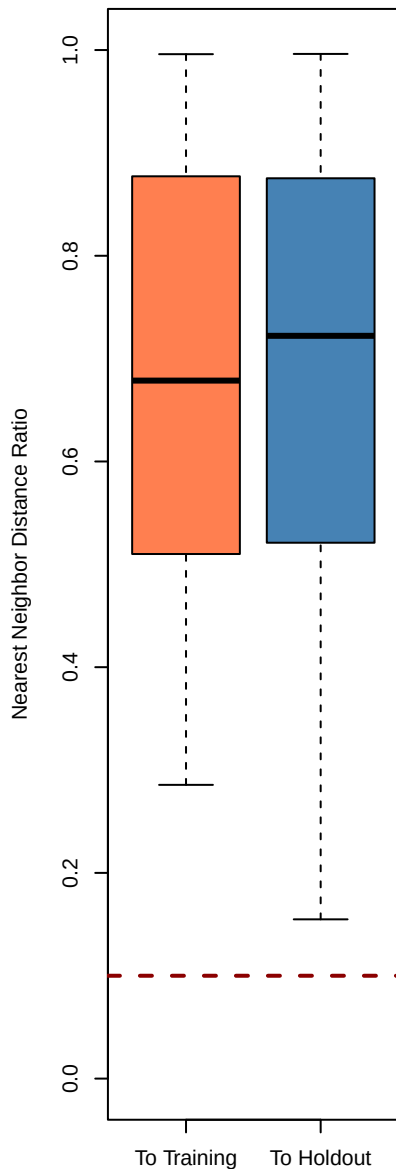
Cumulative Distribution of NNDR



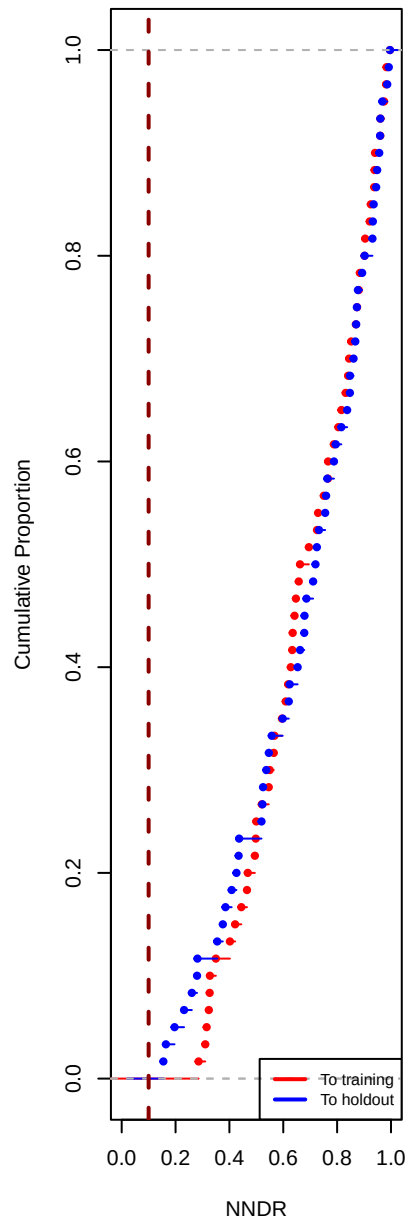
NNDR Distribution



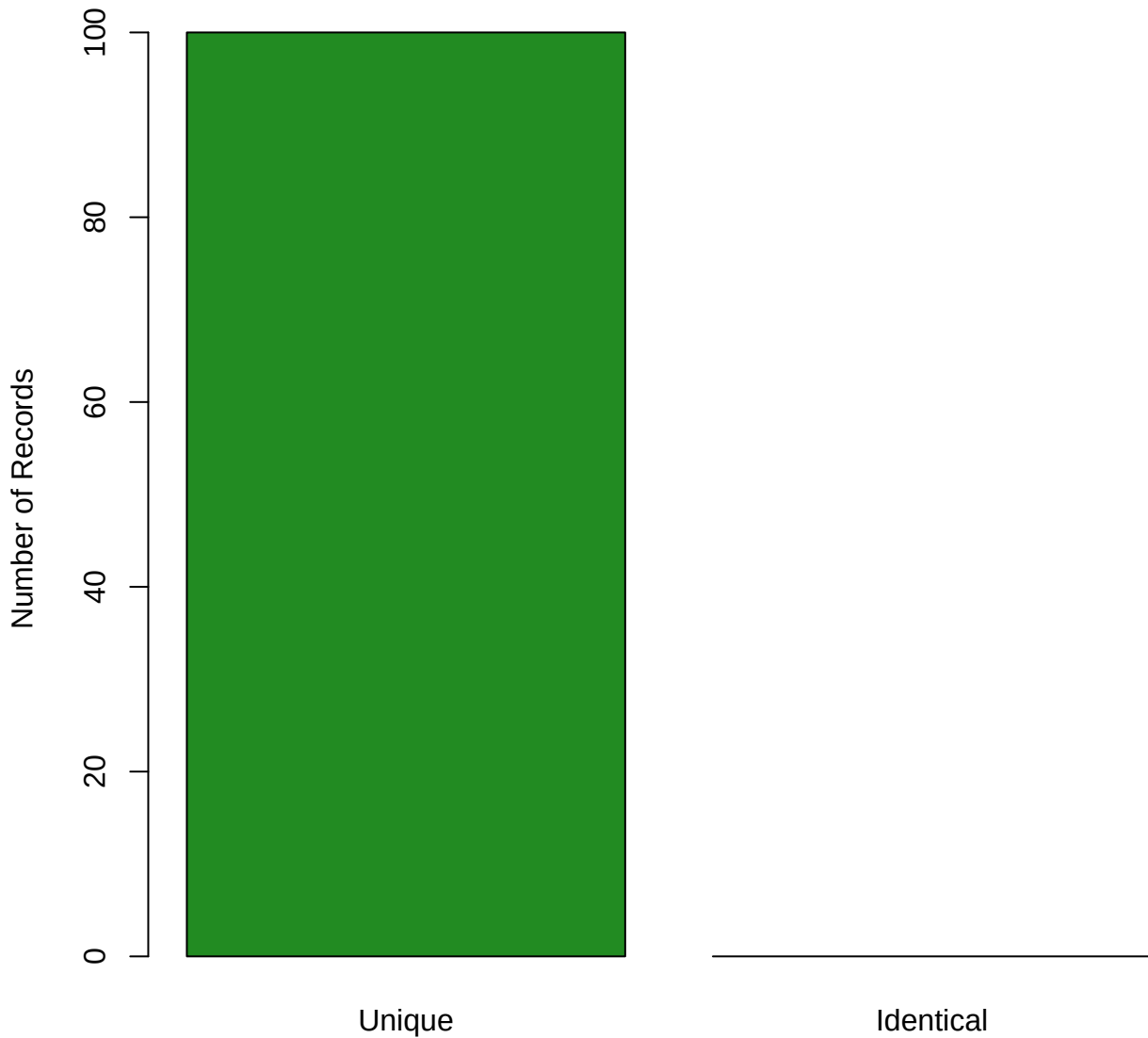
NNDR Comparison



Cumulative Distribution of NNDR



Synthetic Record Status

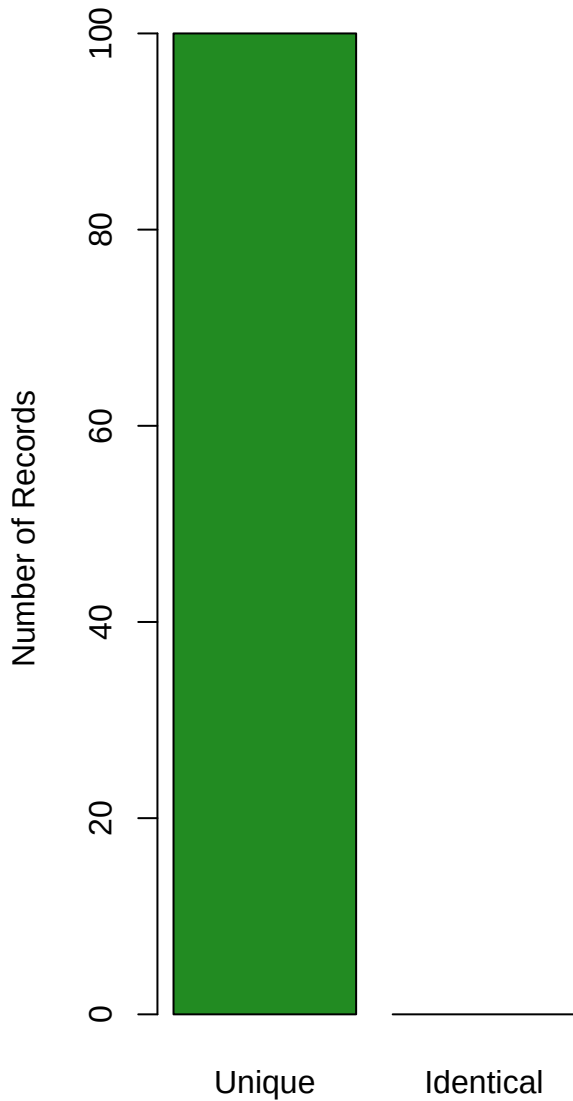


No Identical Records Found

No identical matches to analyze

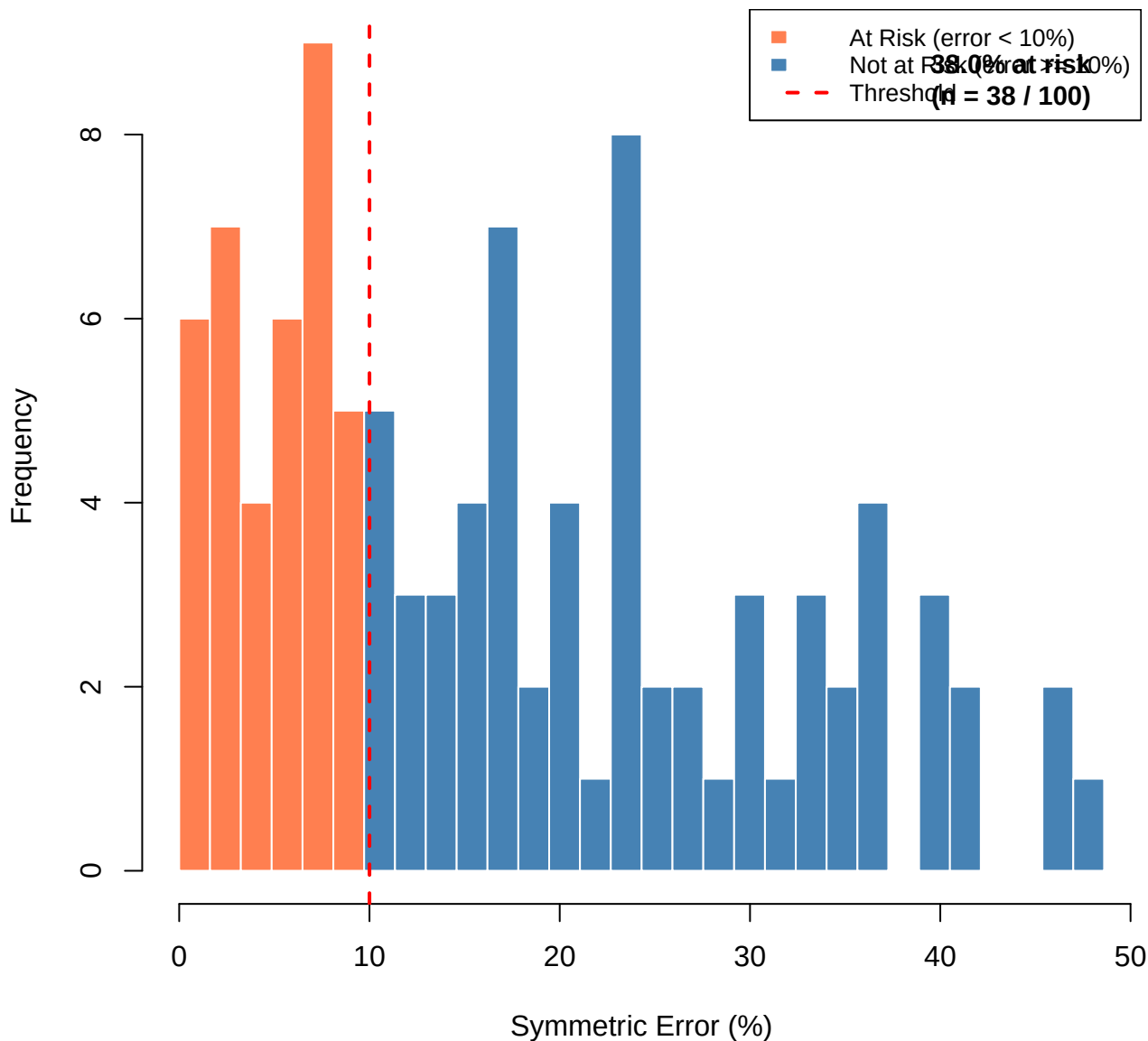
Synthetic Record Status

No Identical Records Found

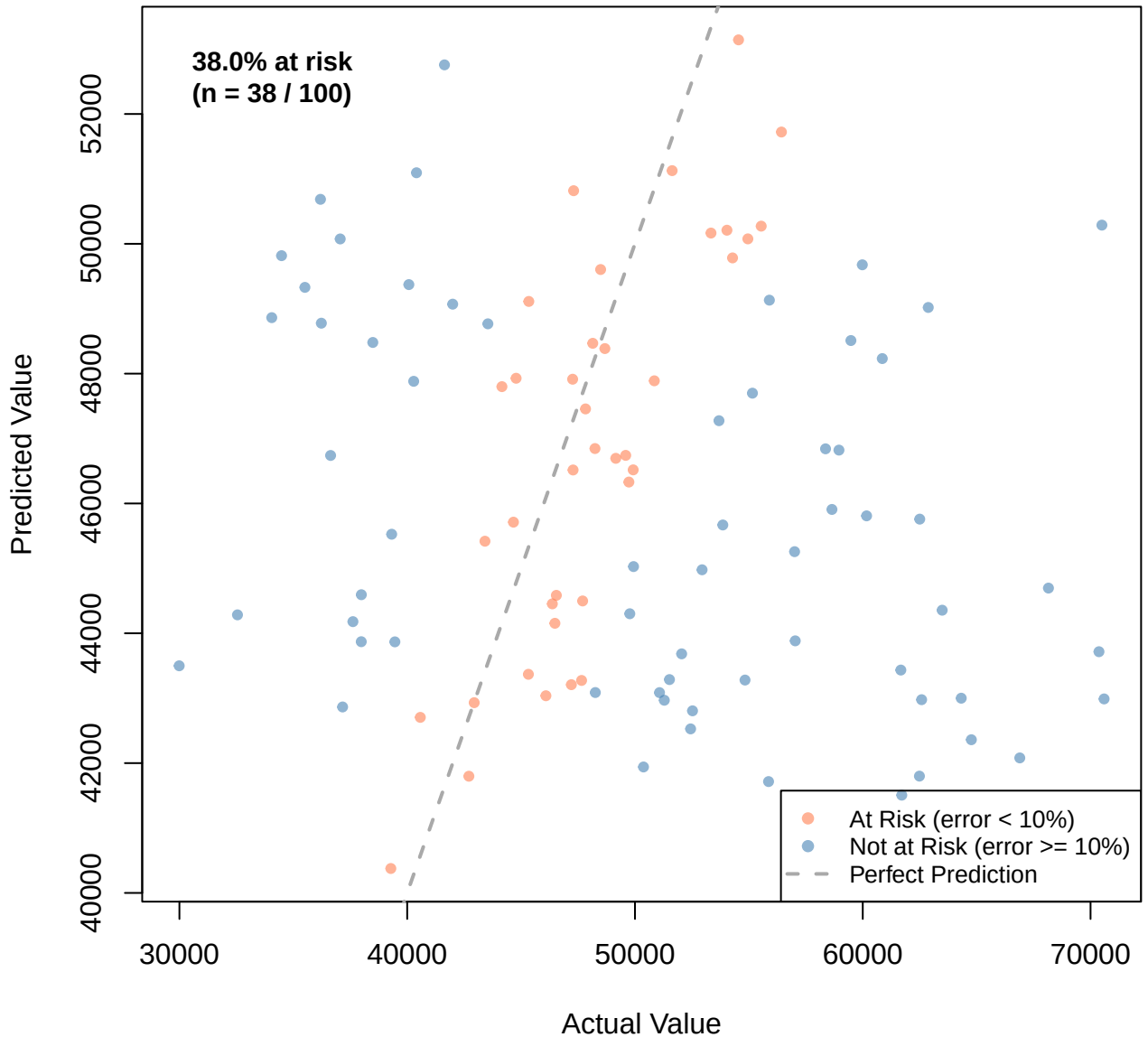


identical matches to analyze

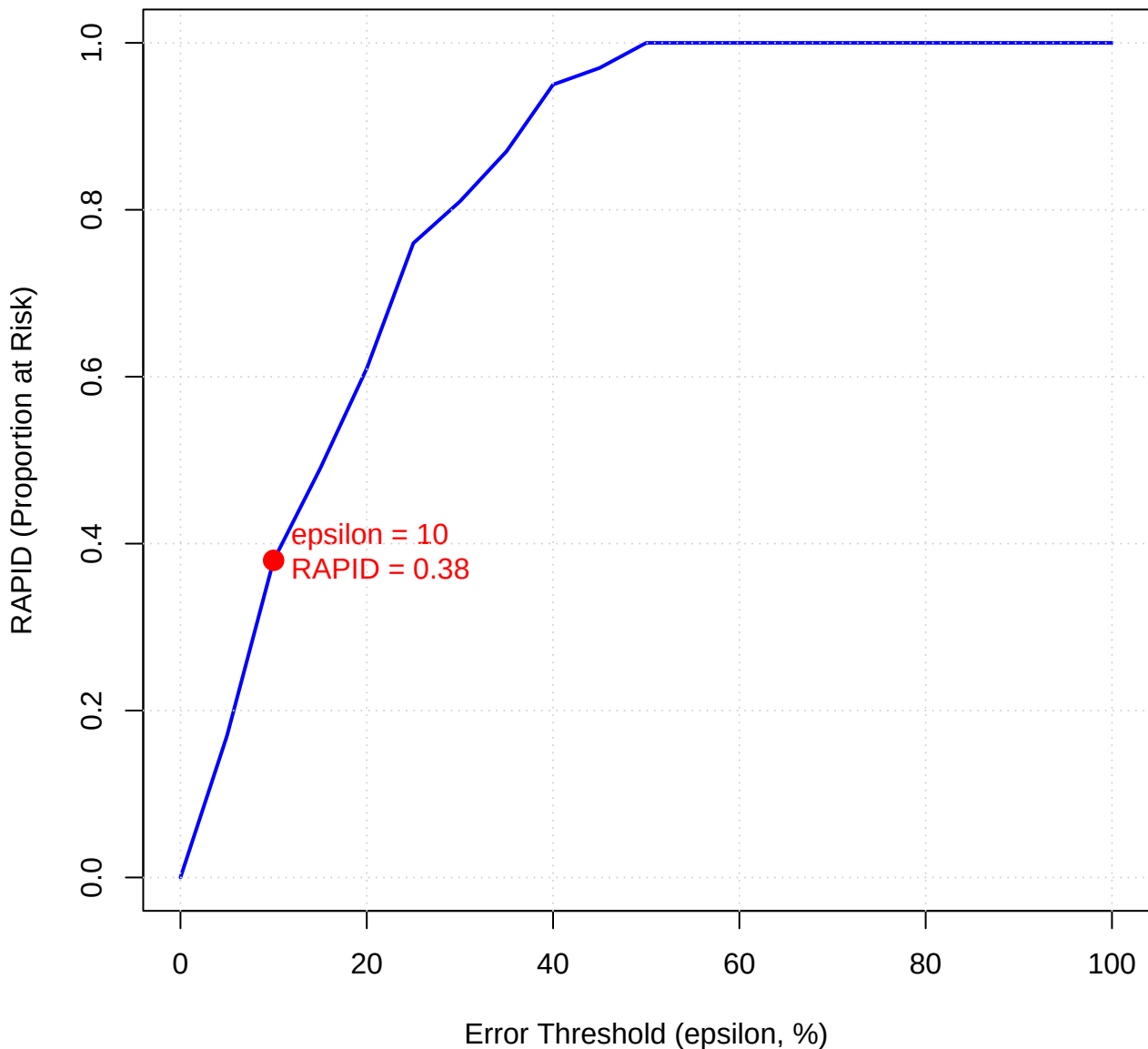
Prediction Error Distribution



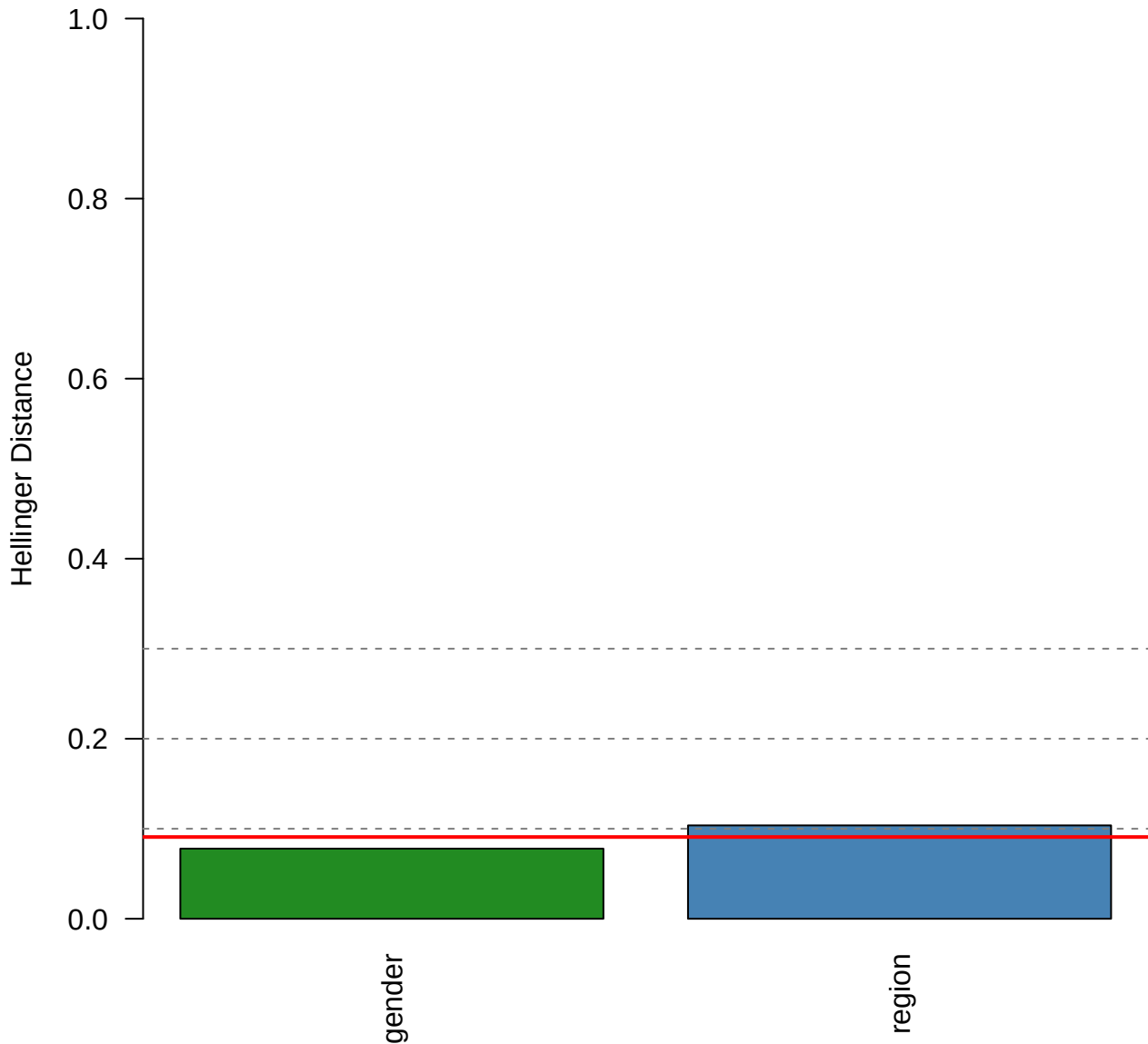
Predicted vs Actual Values



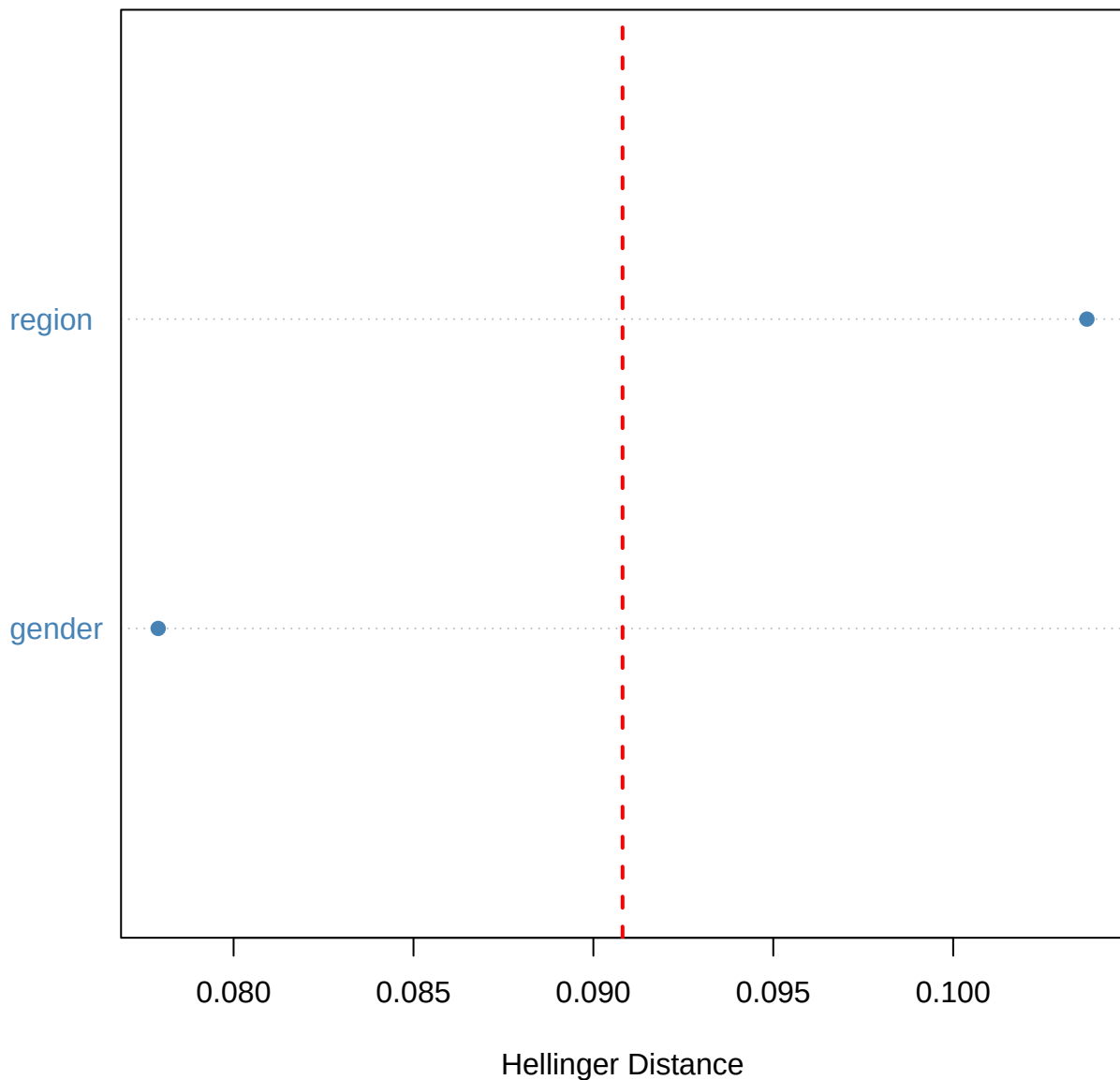
Threshold Sensitivity Curve



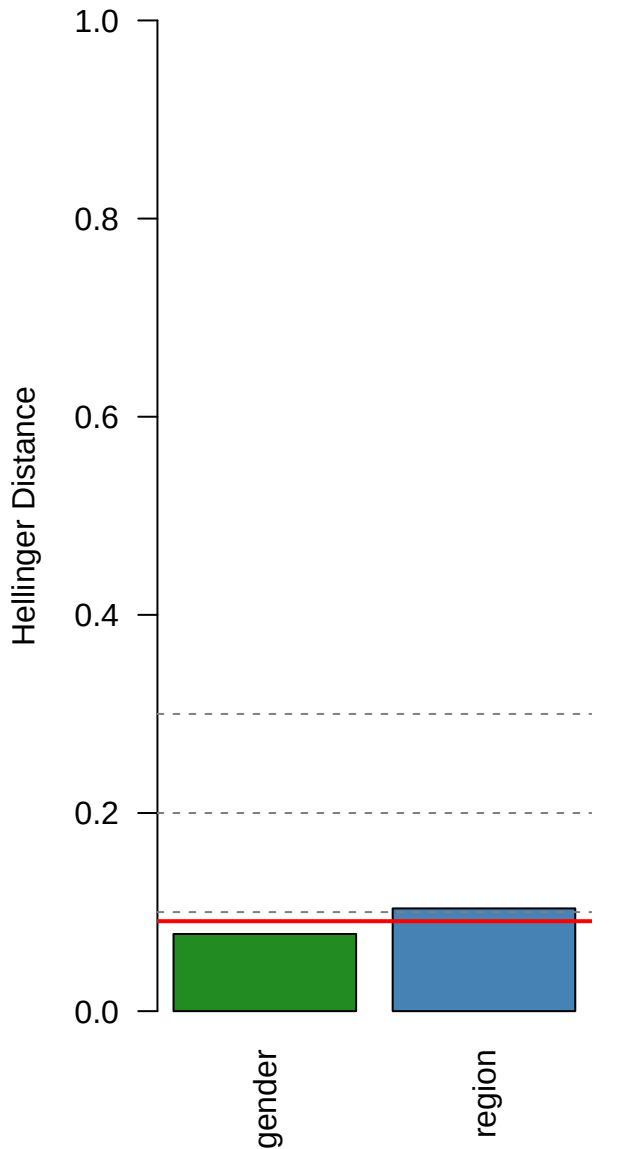
Hellinger Distance by Variable



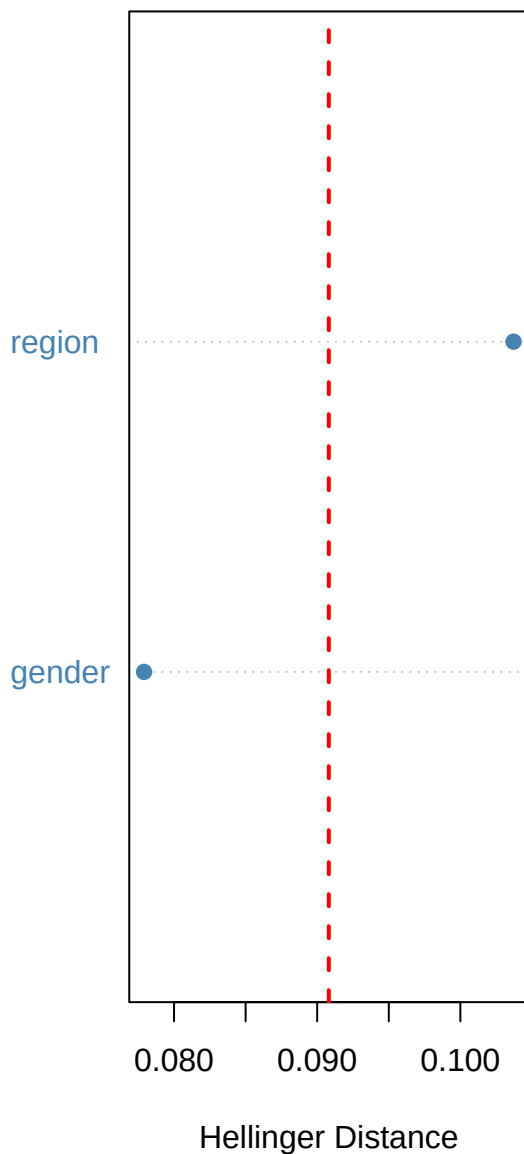
Hellinger Distance (sorted)



Hellinger Distance by Variable

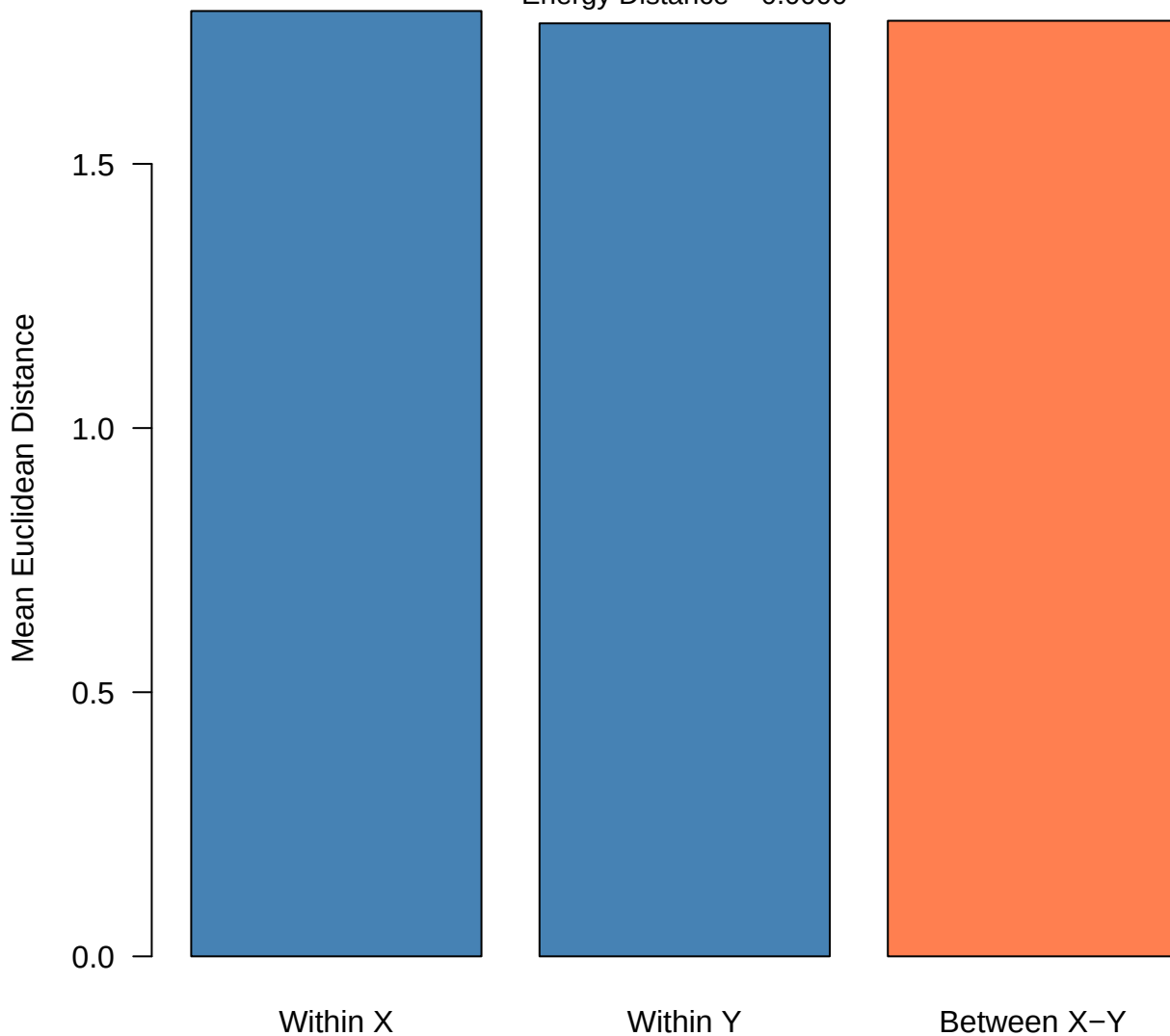


Hellinger Distance (sorted)



Distance Components

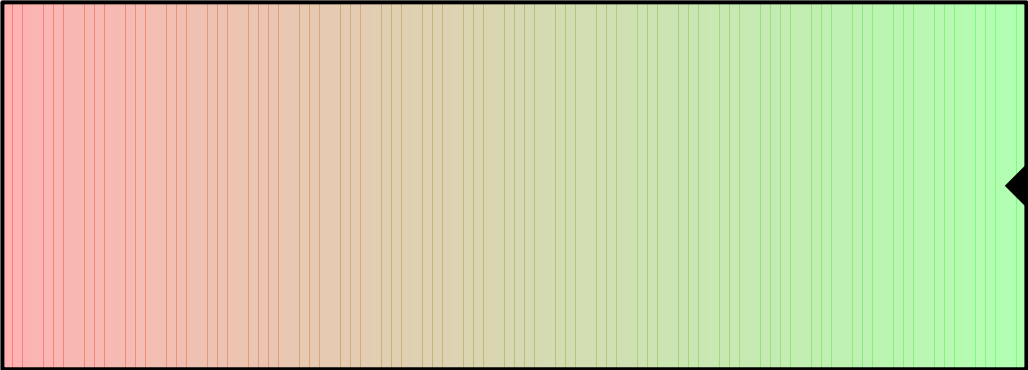
Energy Distance = 0.0000



Utility Score

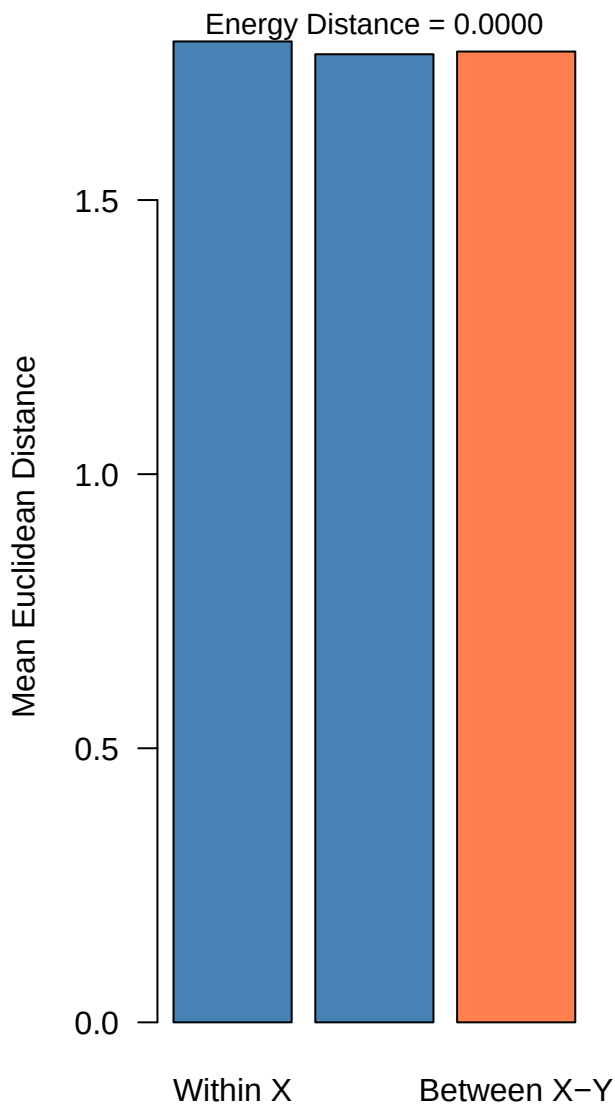
Poor

Good

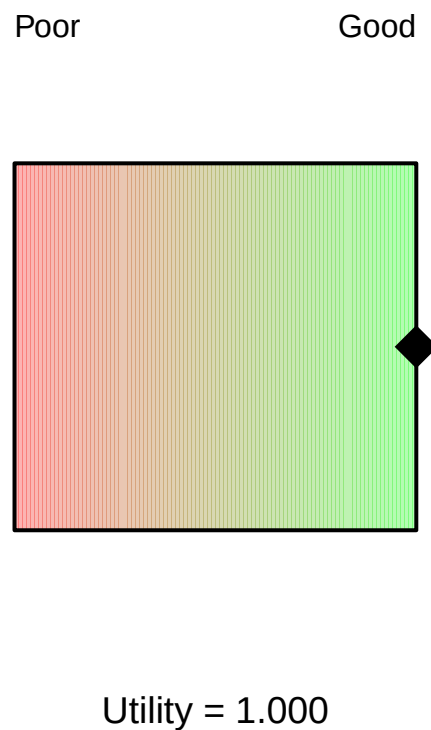


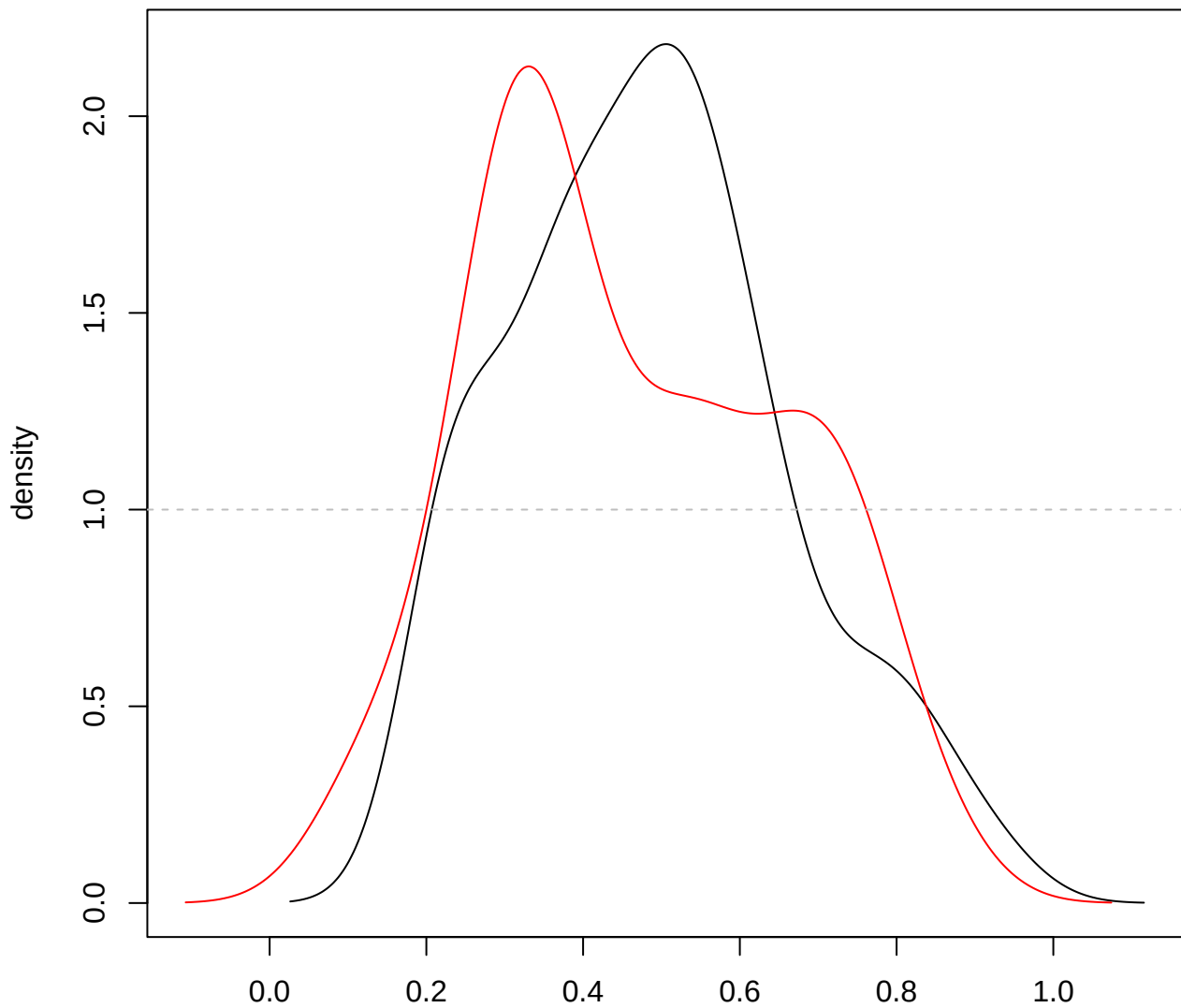
Utility = 1.000

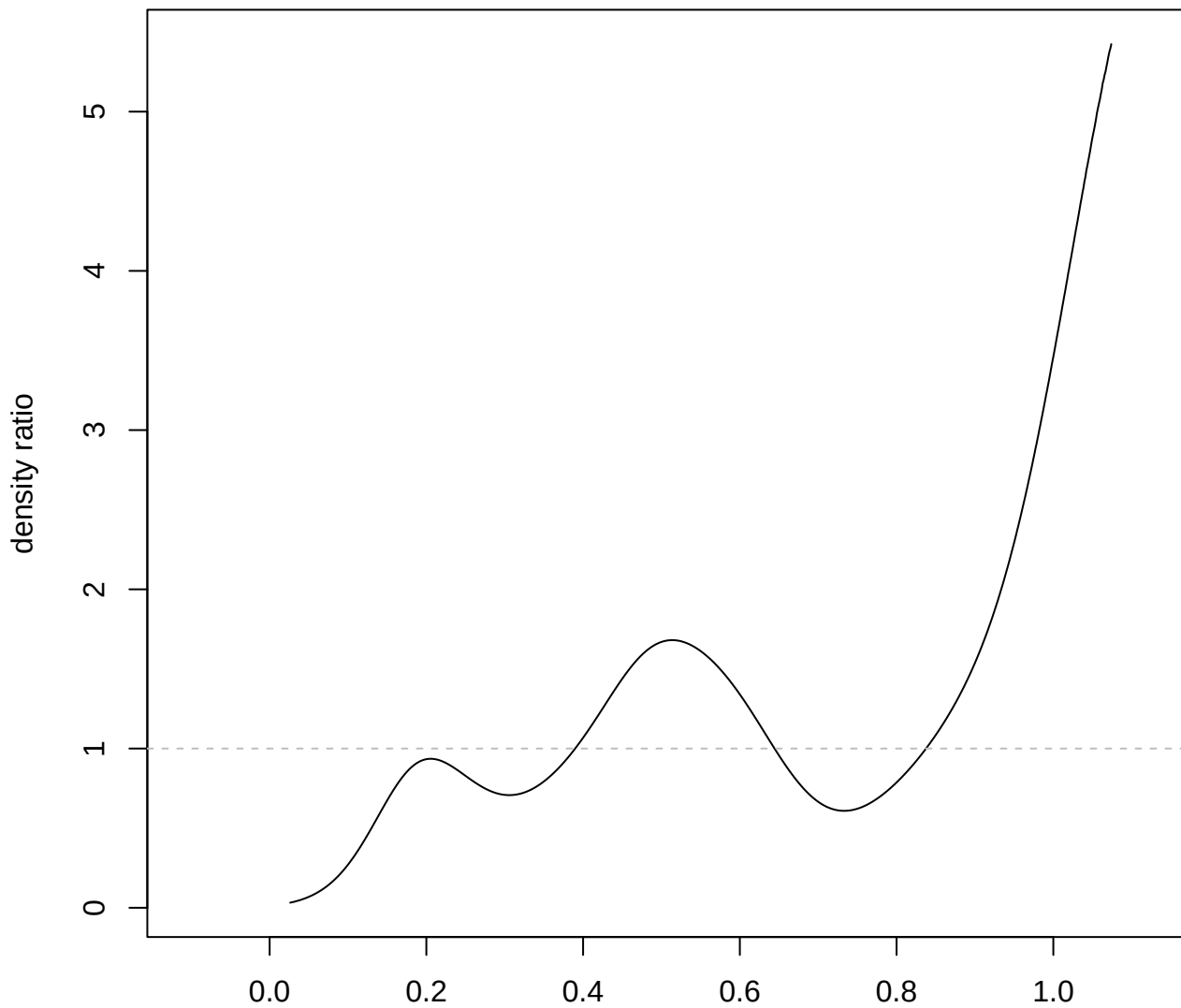
Distance Components

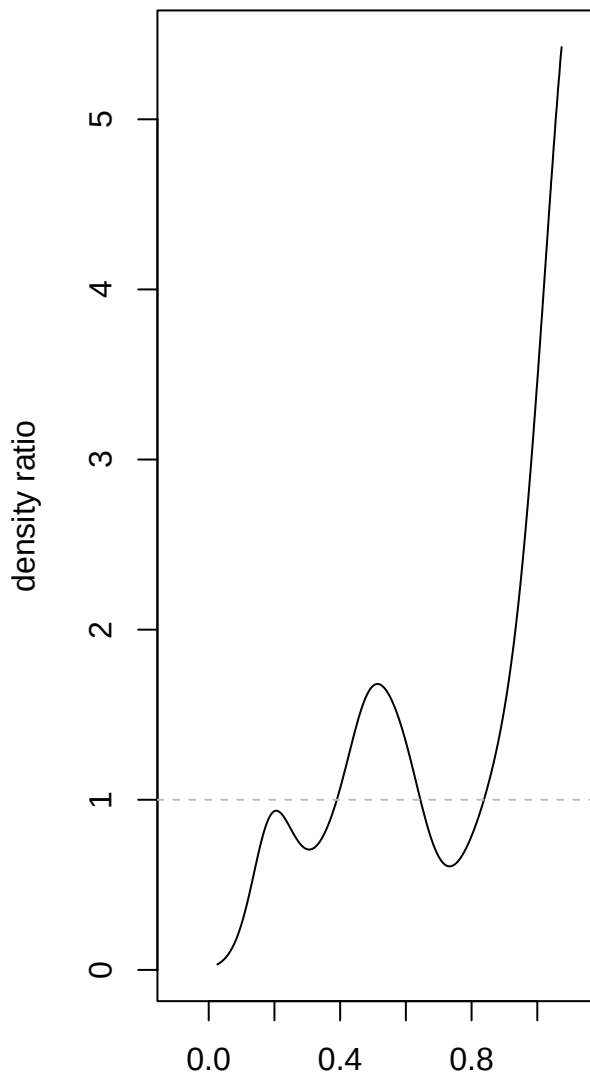
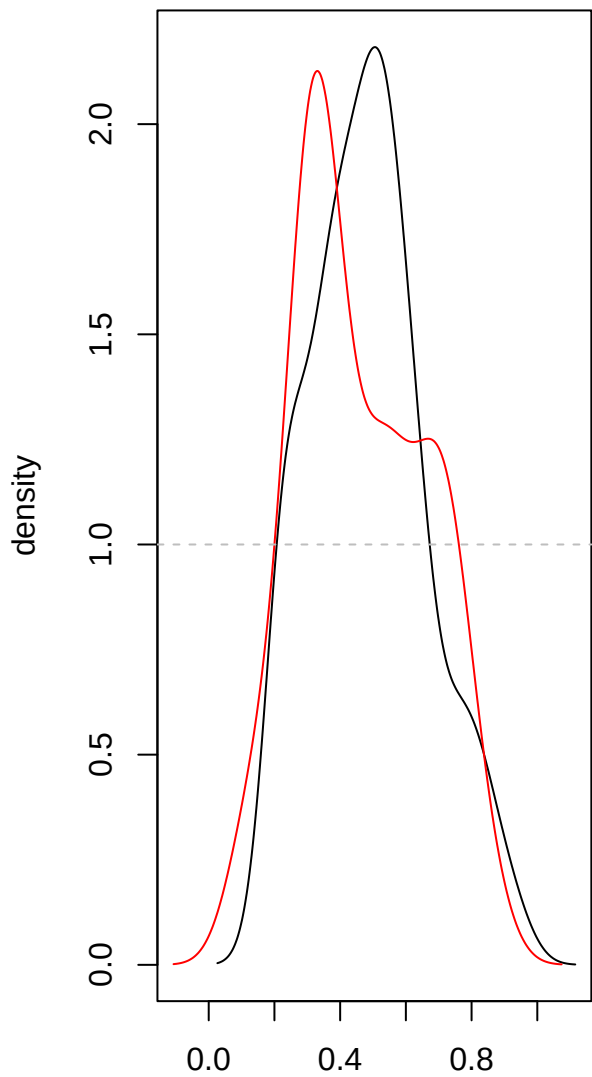


Utility Score

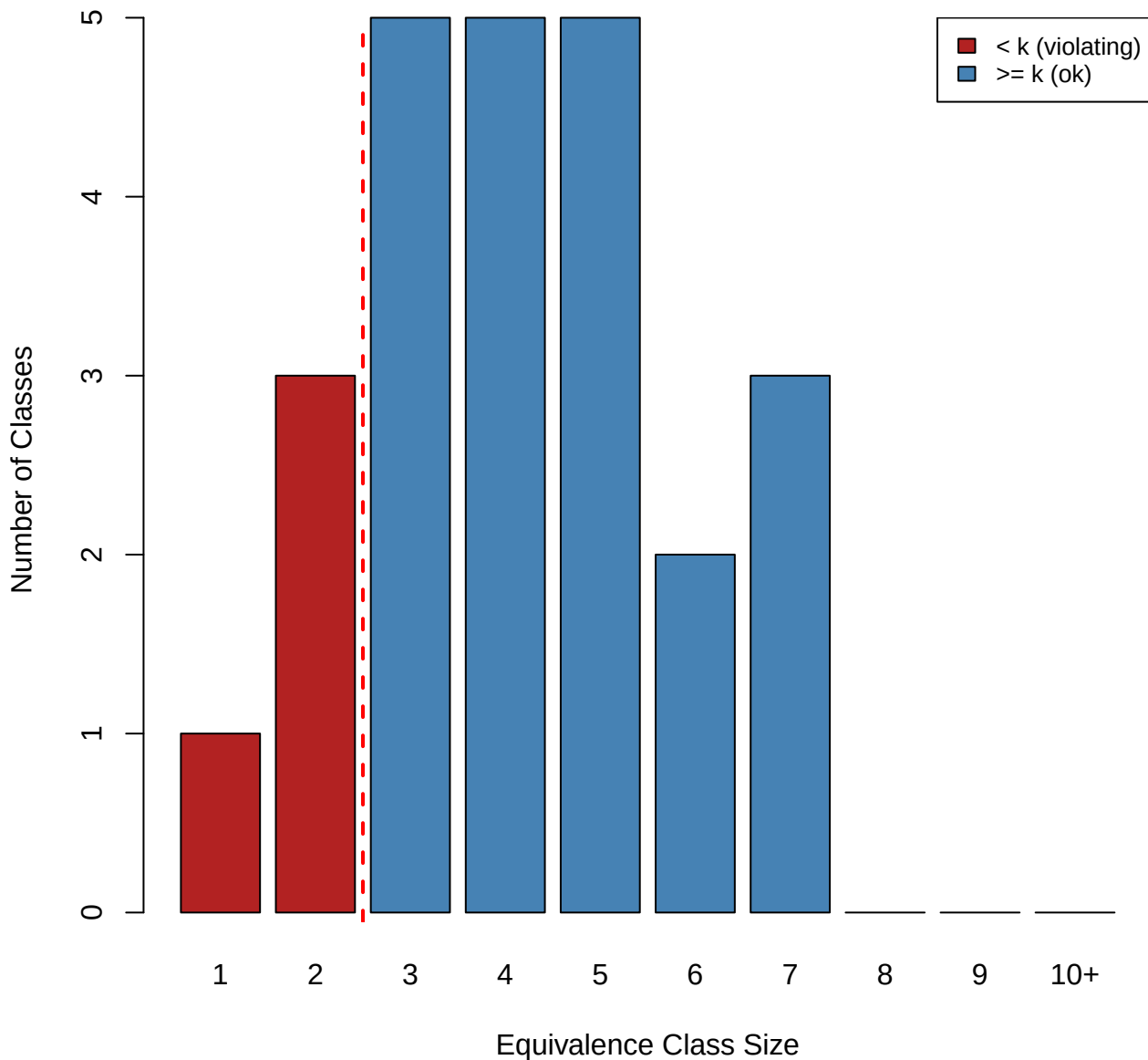




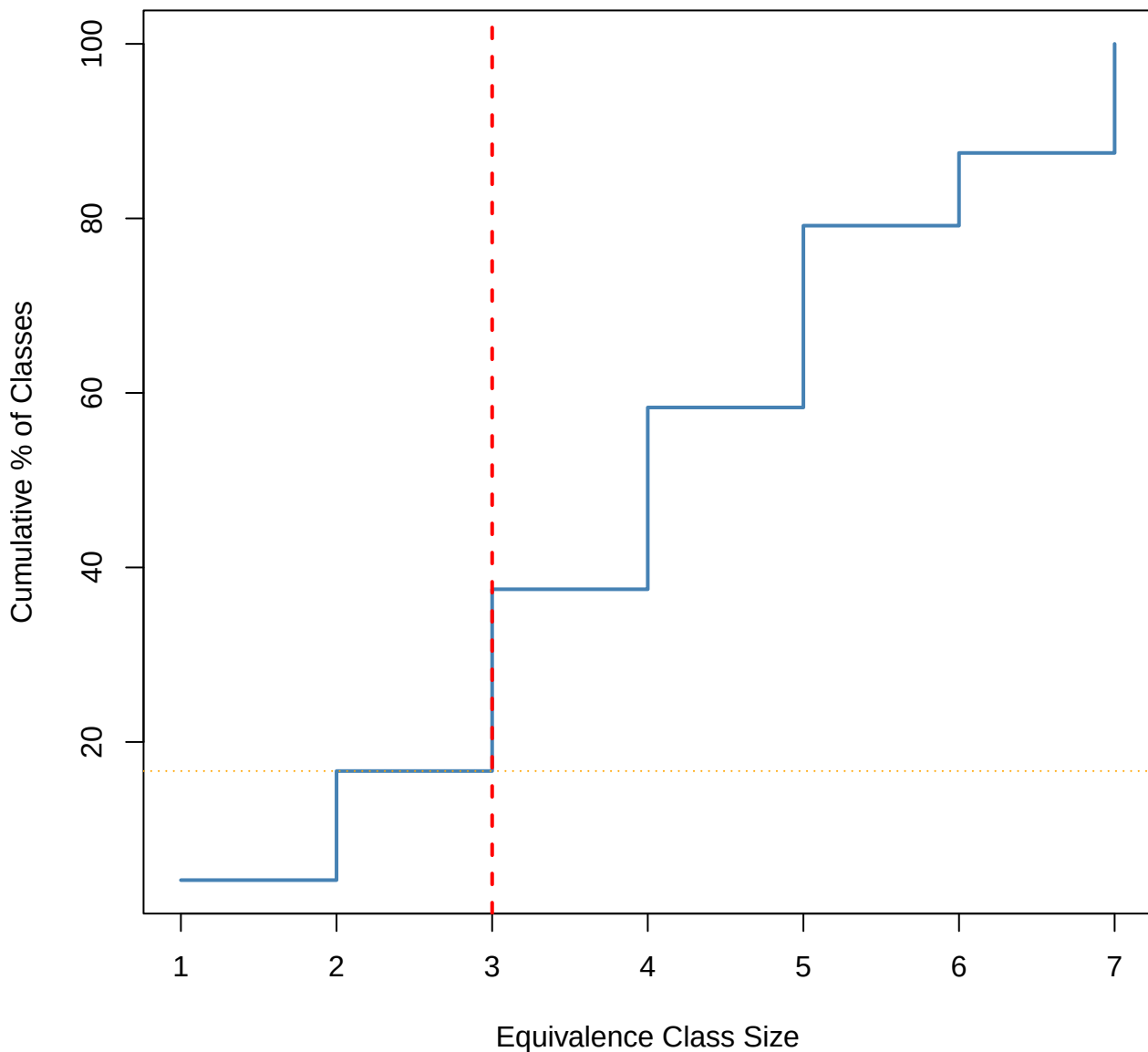




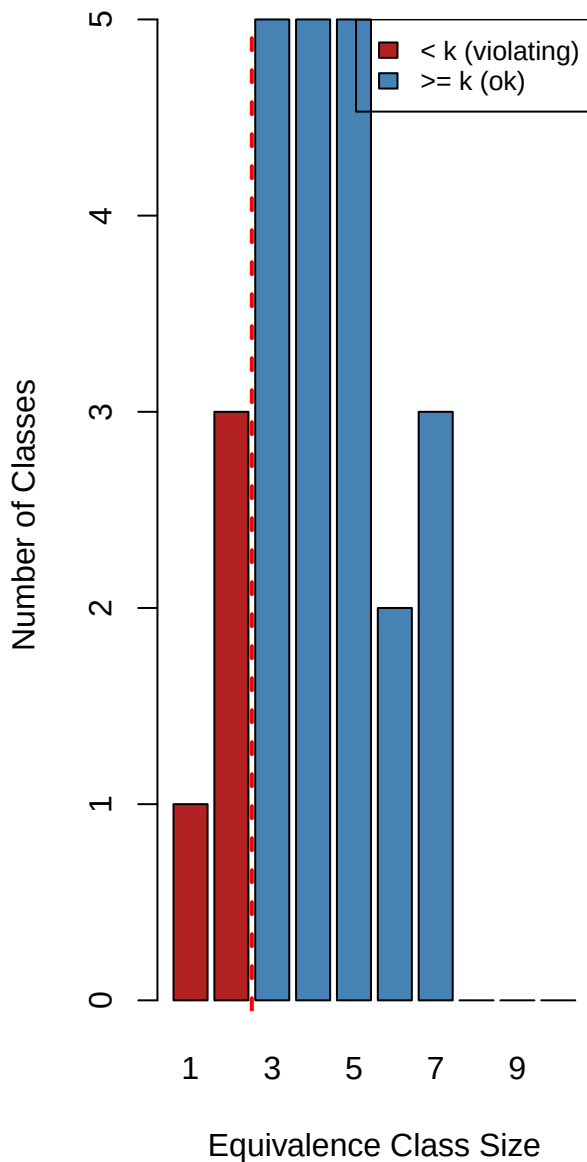
Equivalence Class Sizes ($k = 3$)



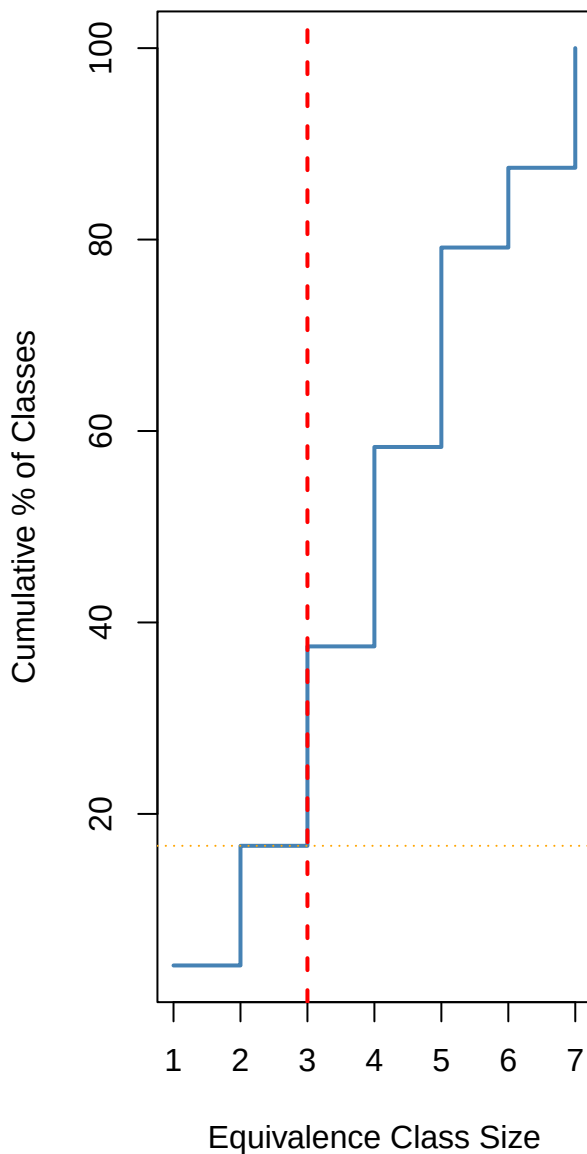
Cumulative Distribution of EC Sizes



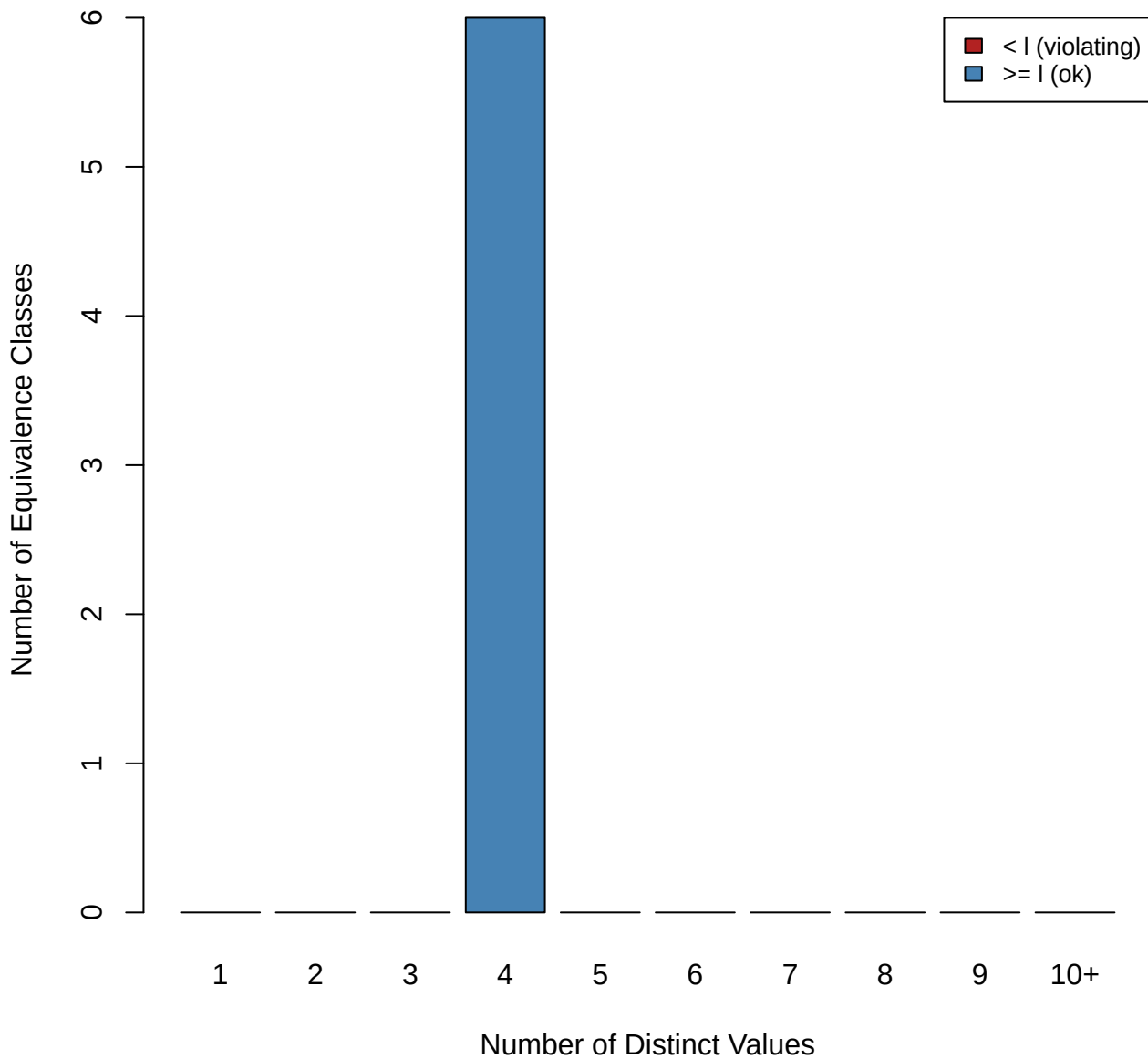
Equivalence Class Sizes ($k = 3$)



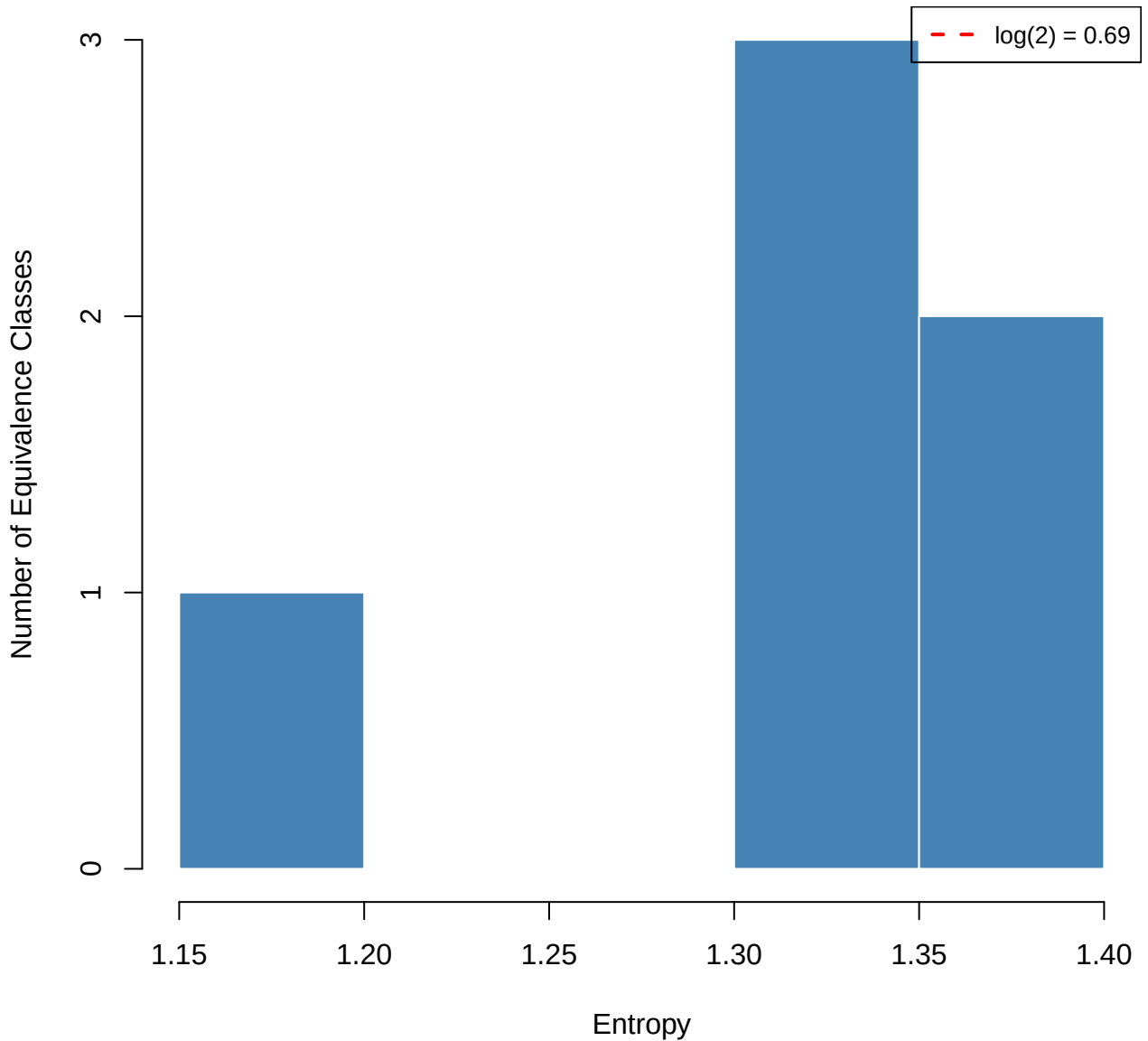
Cumulative Distribution of EC Size



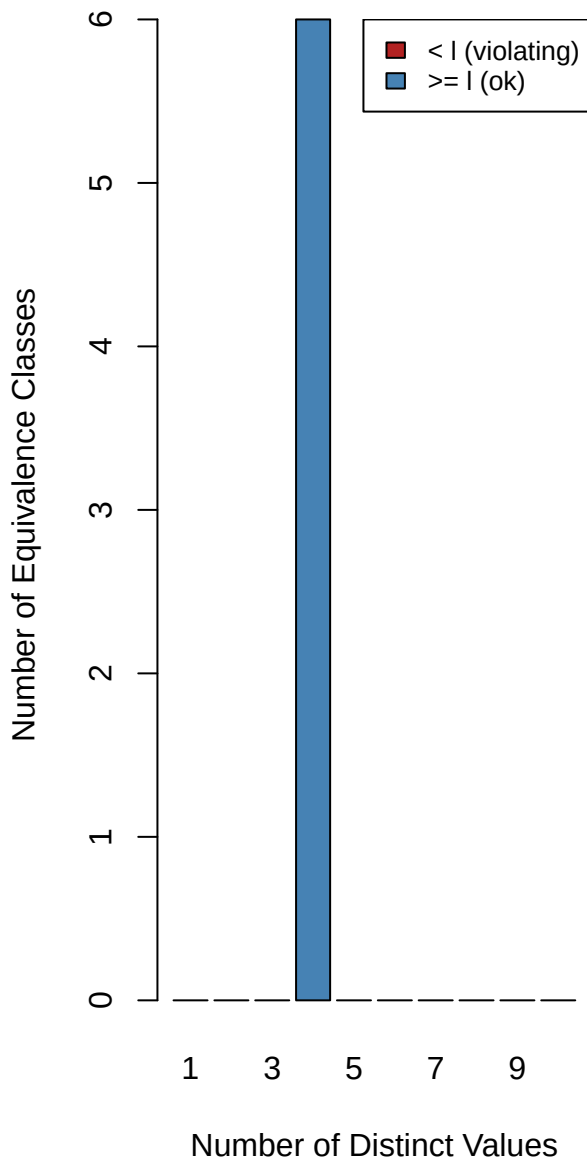
Distinct Sensitive Values per EC (I = 2)



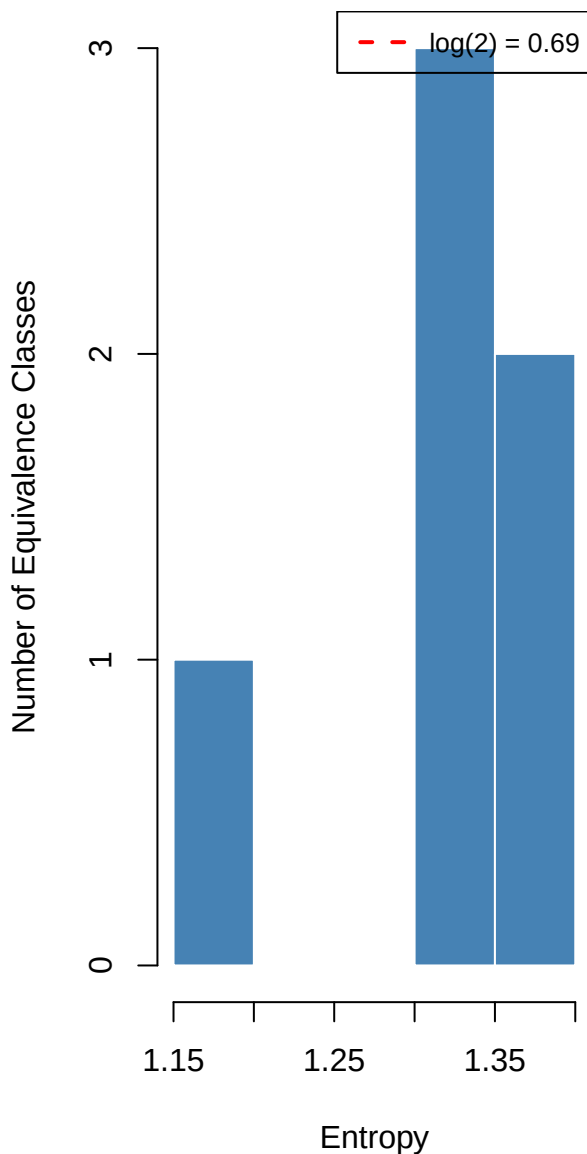
Entropy Distribution (l = 2)



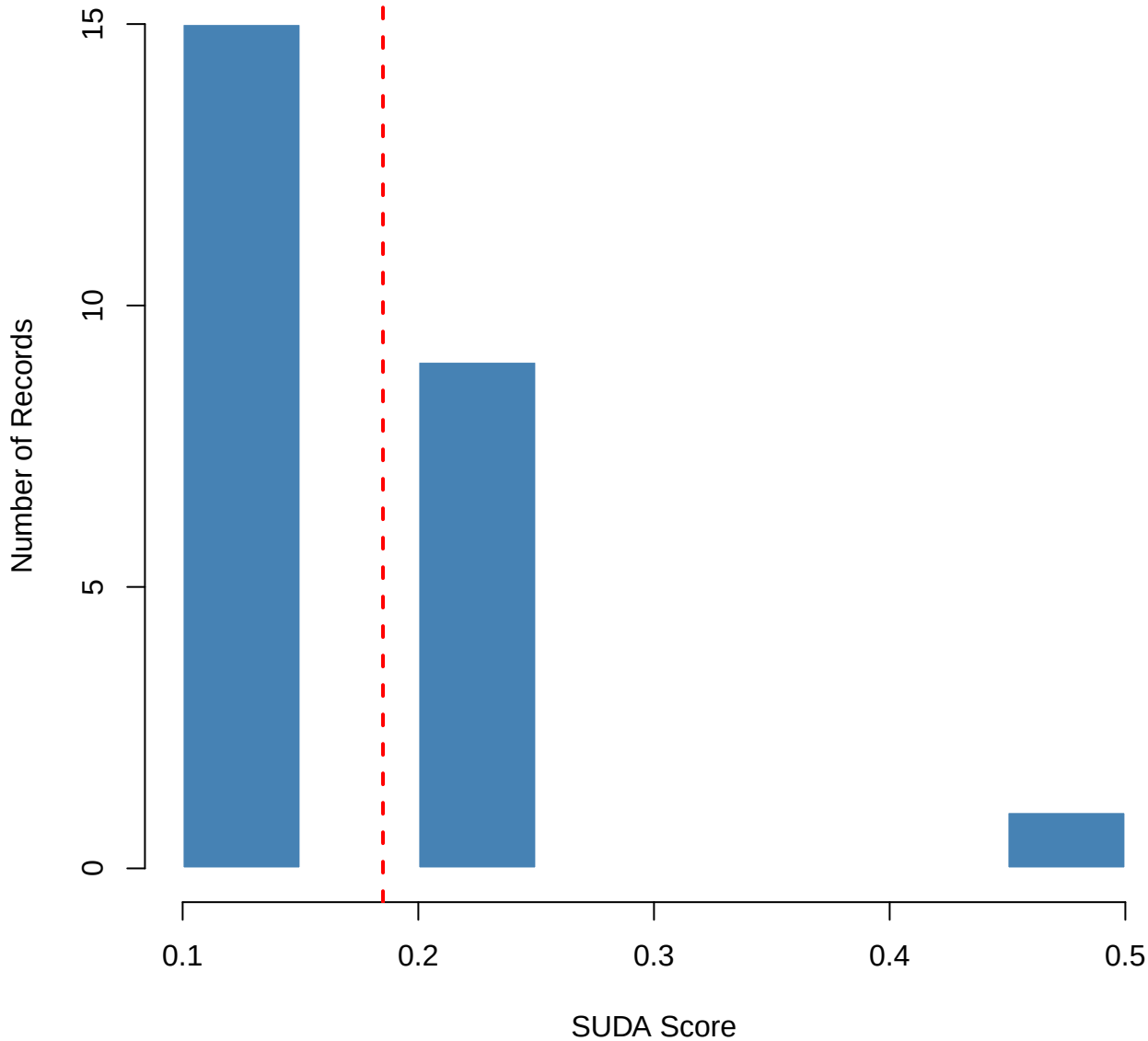
Distinct Sensitive Values per EC (l =



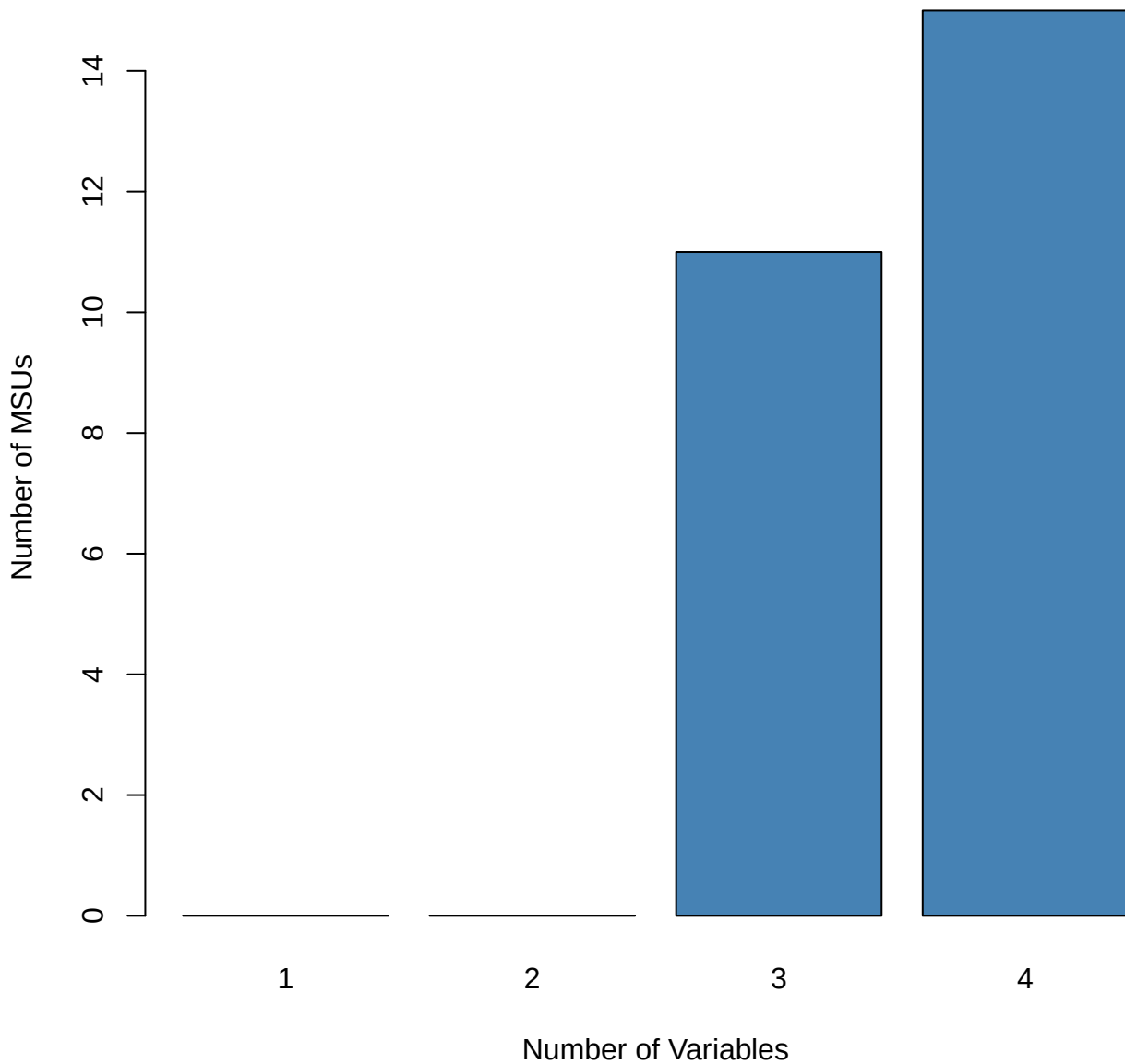
Entropy Distribution (l = 2)



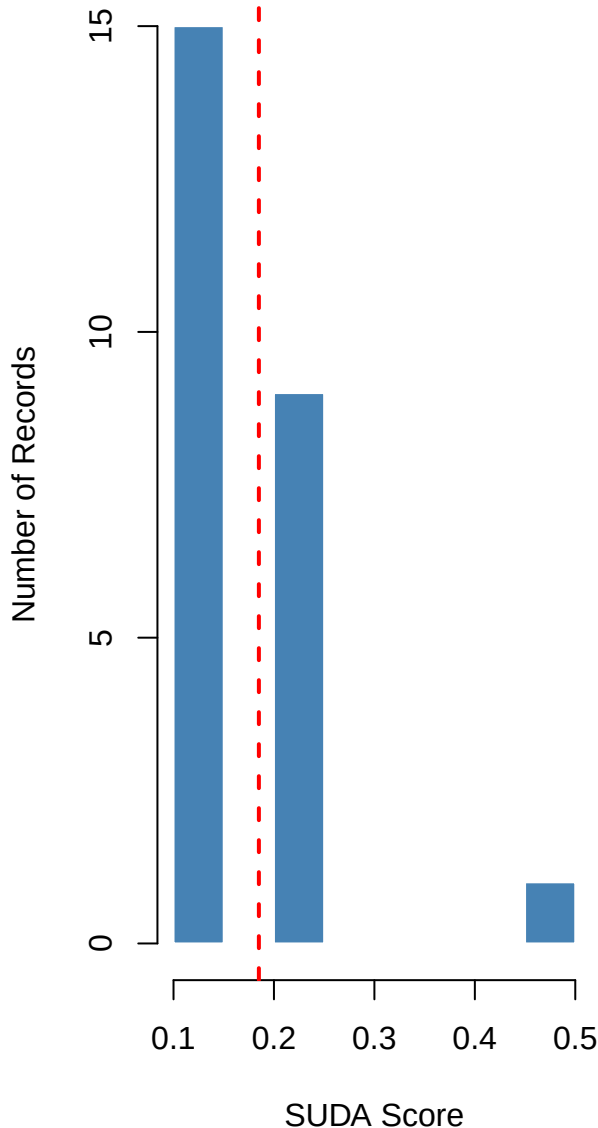
Distribution of SUDA Scores (non-zero)



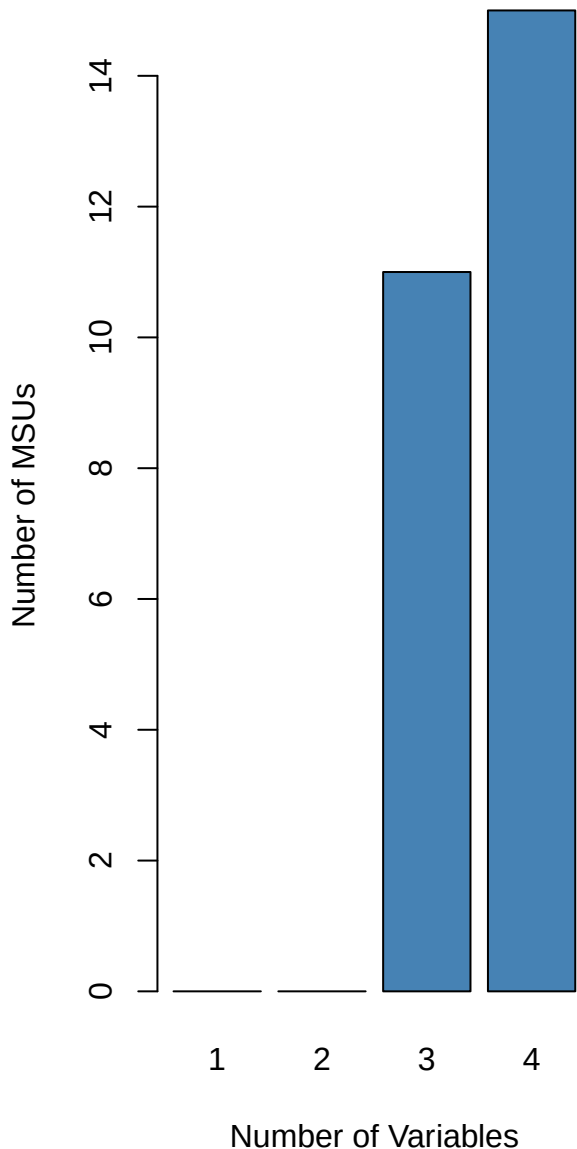
MSUs by Variable Subset Size



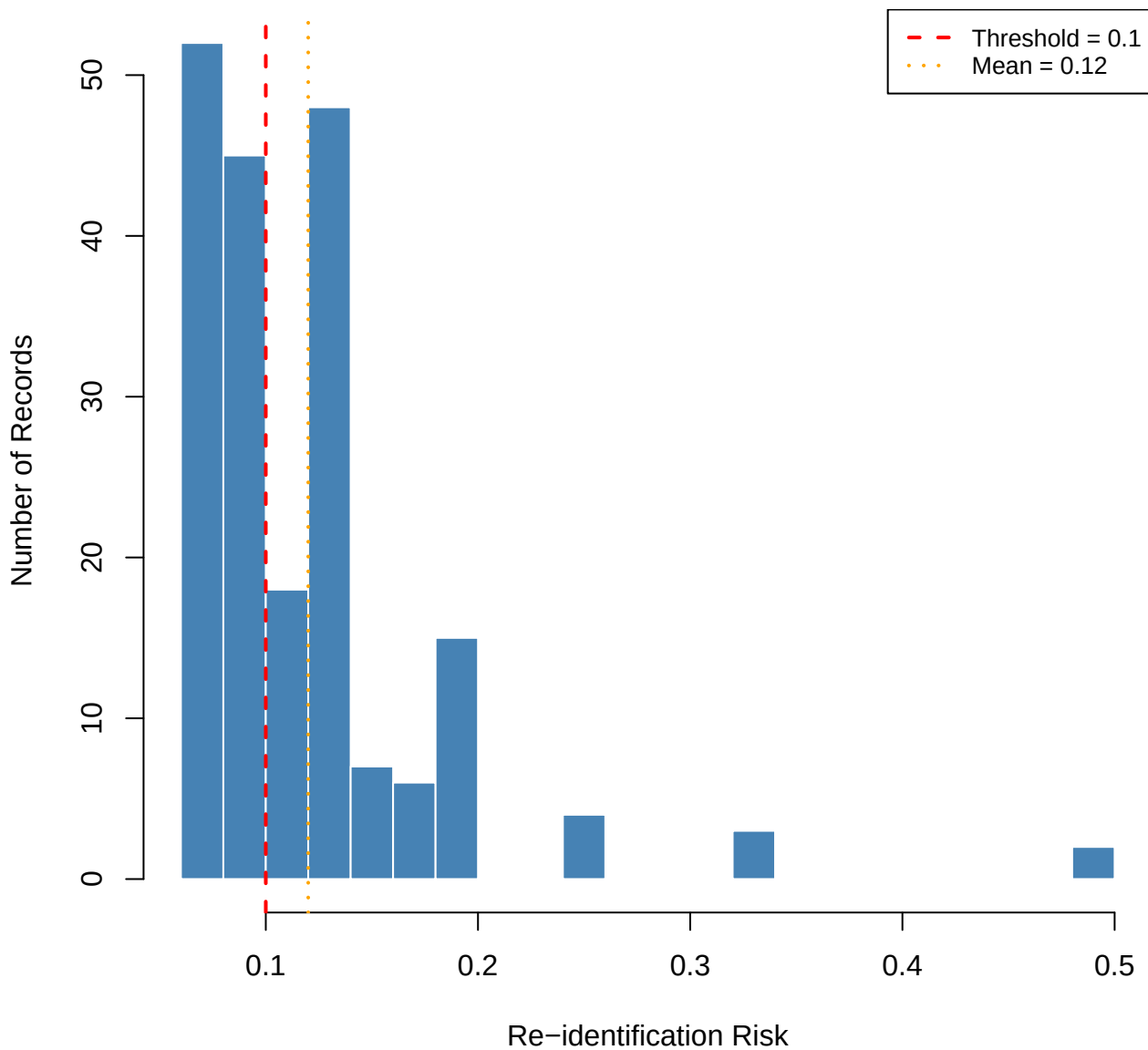
Distribution of SUDA Scores (non-zero)



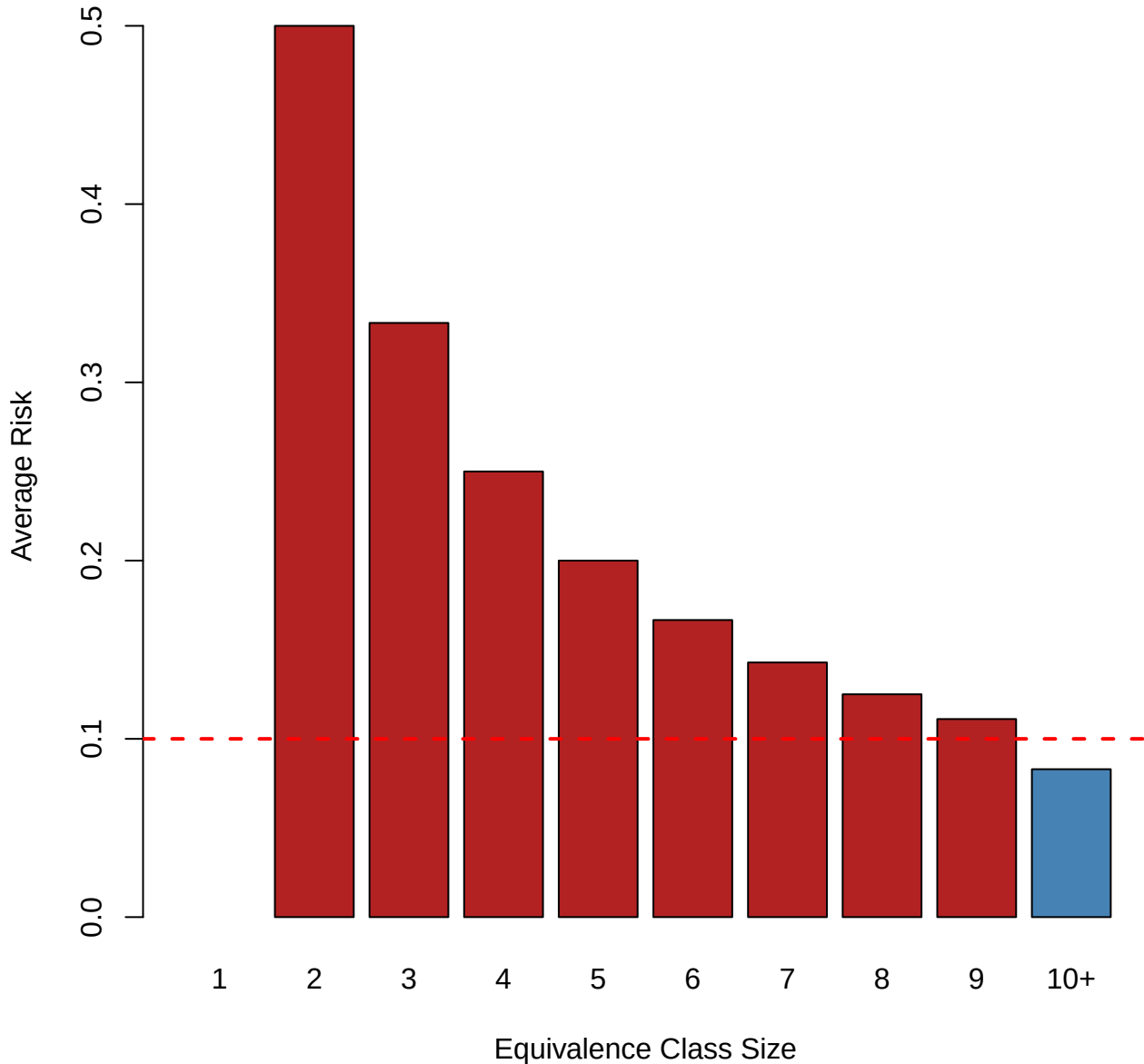
MSUs by Variable Subset Size



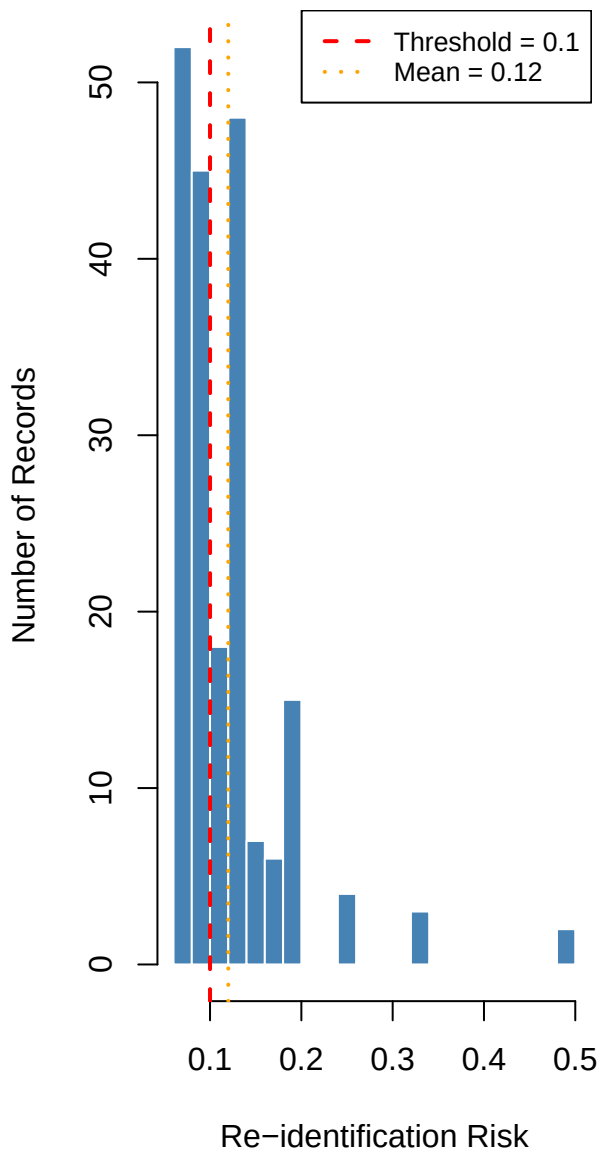
Distribution of Individual Risk



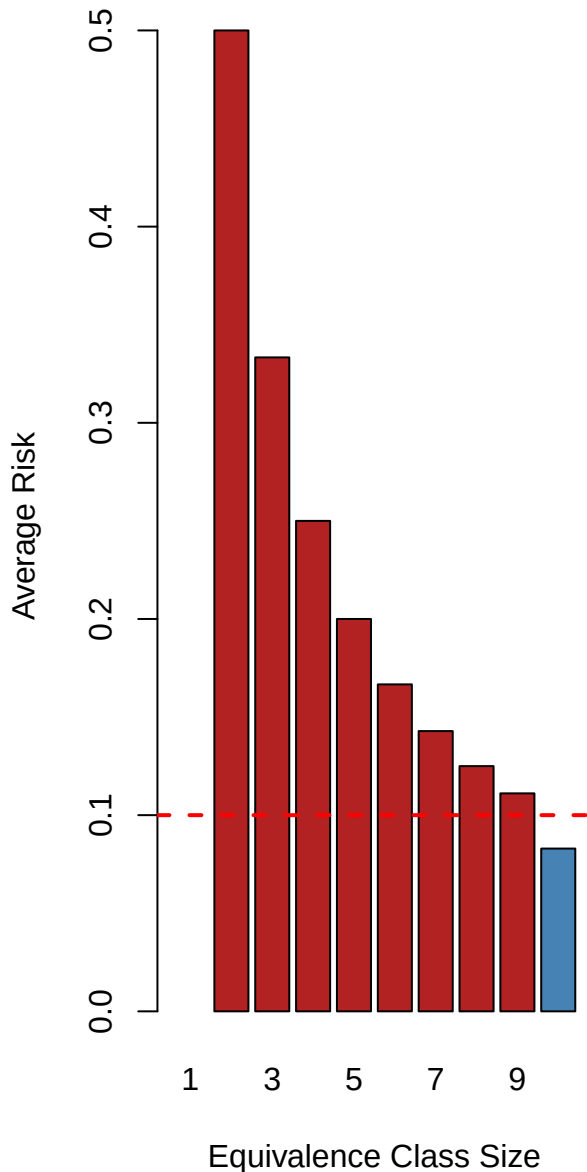
Average Risk by Equivalence Class Size



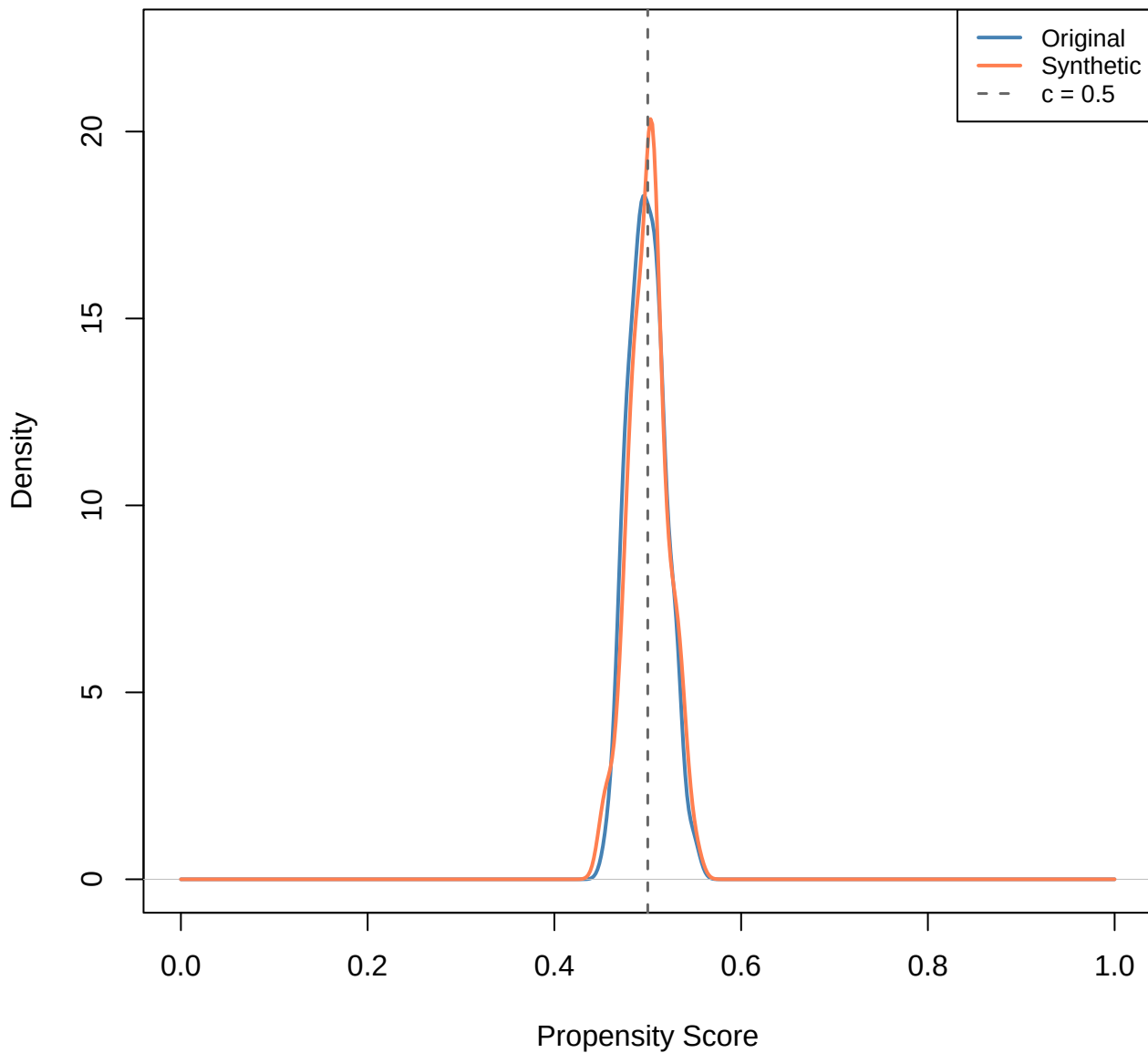
Distribution of Individual Risk



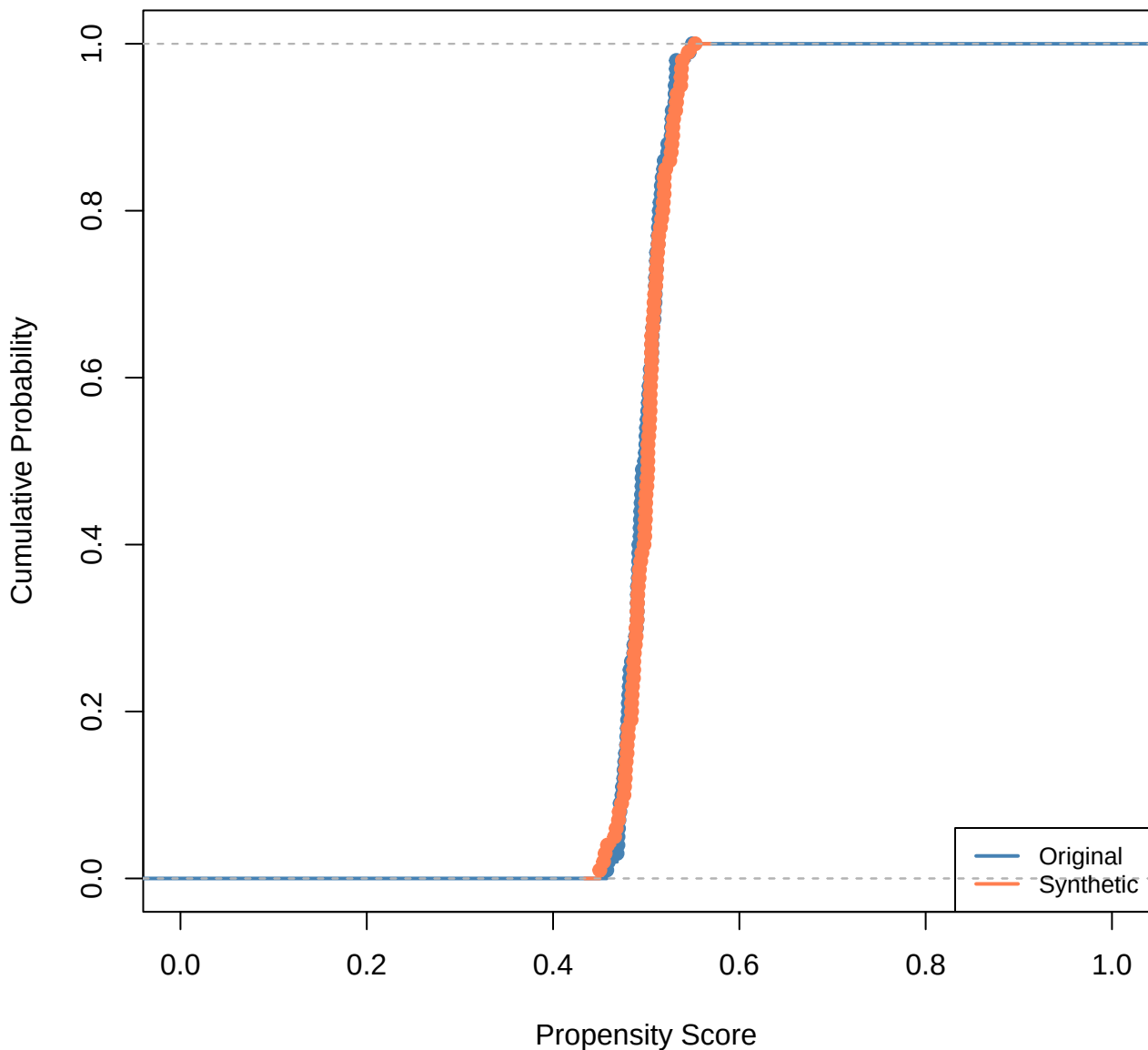
Average Risk by Equivalence Class Size



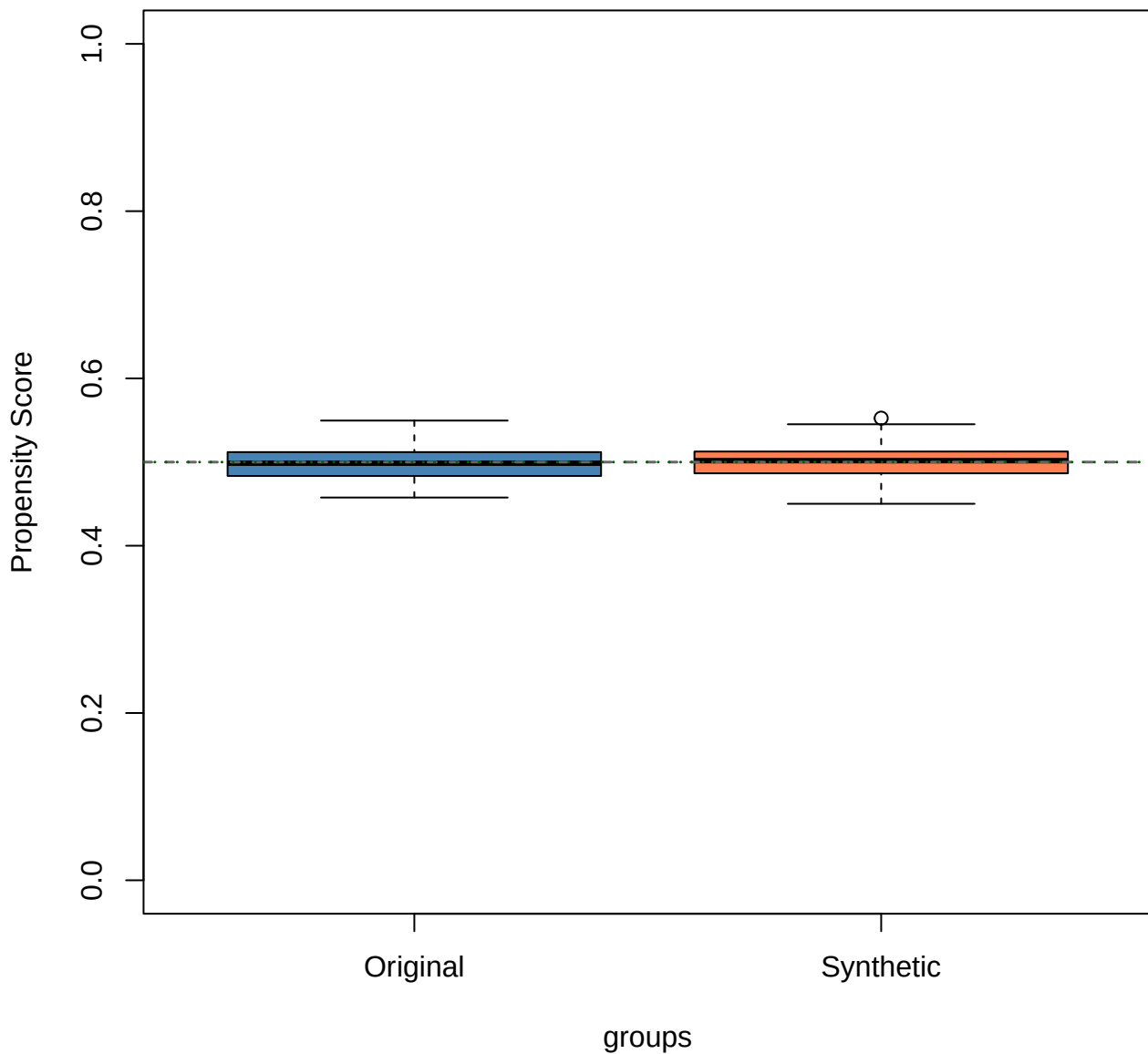
Propensity Score Densities



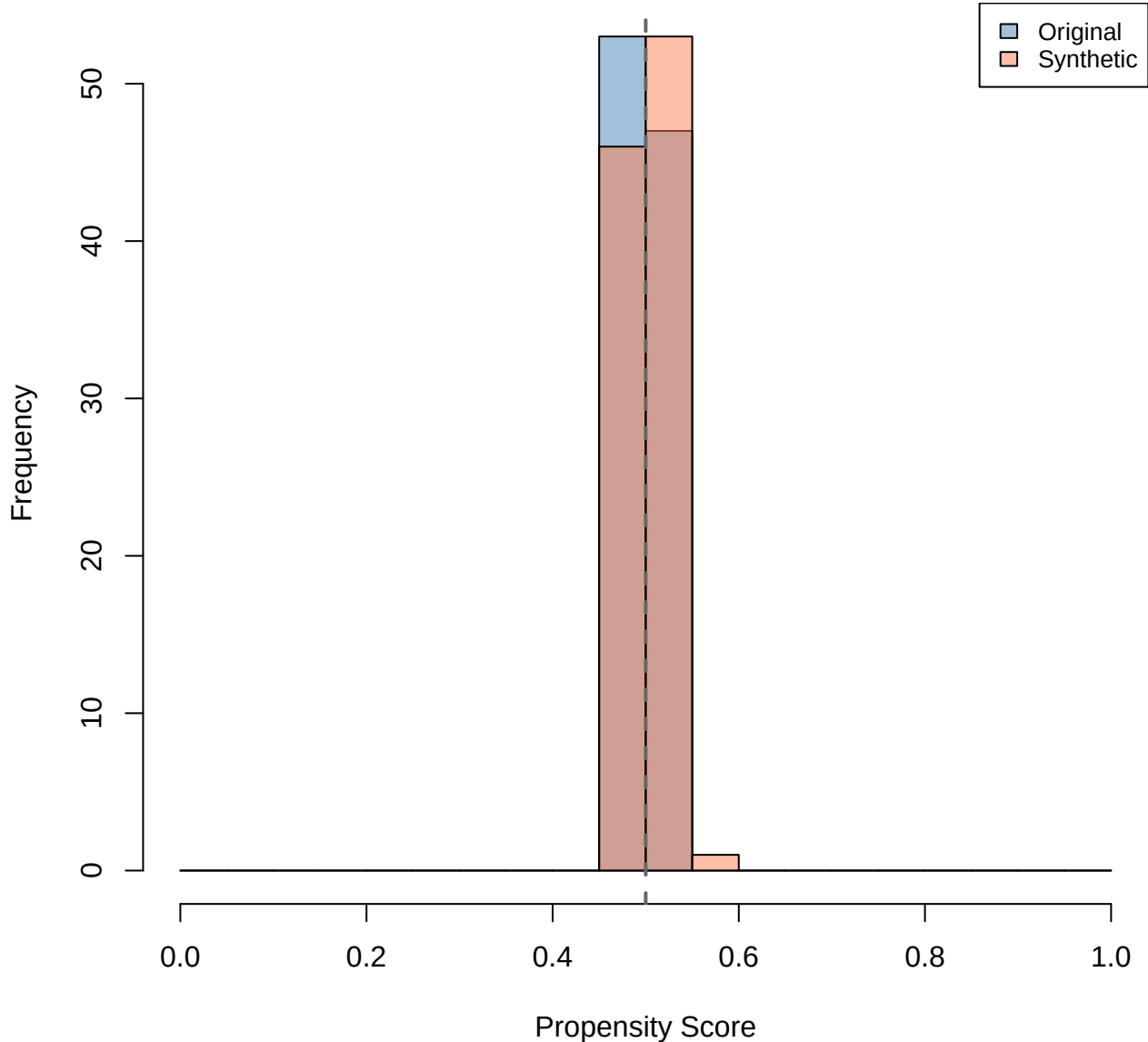
ECDF (SPECKS = 0.1)



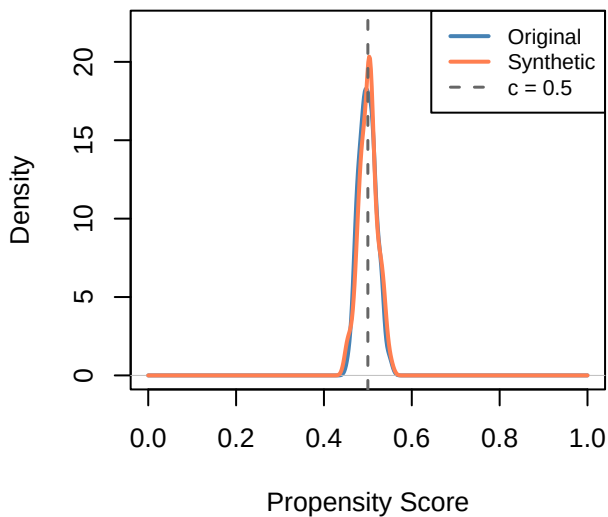
Propensity Scores (PO50 = 53.5%)



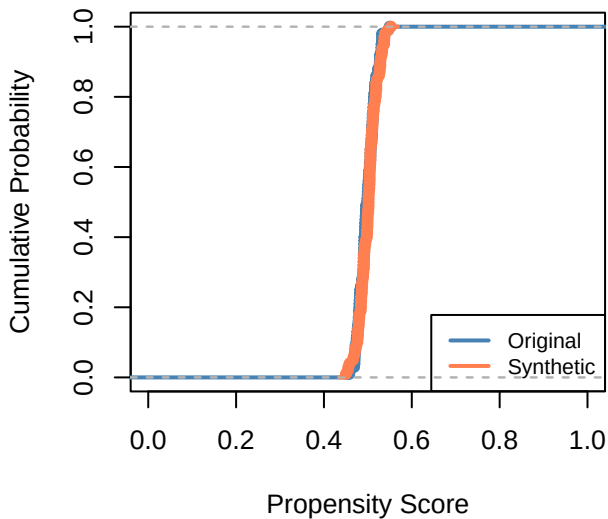
Propensity Score Distributions



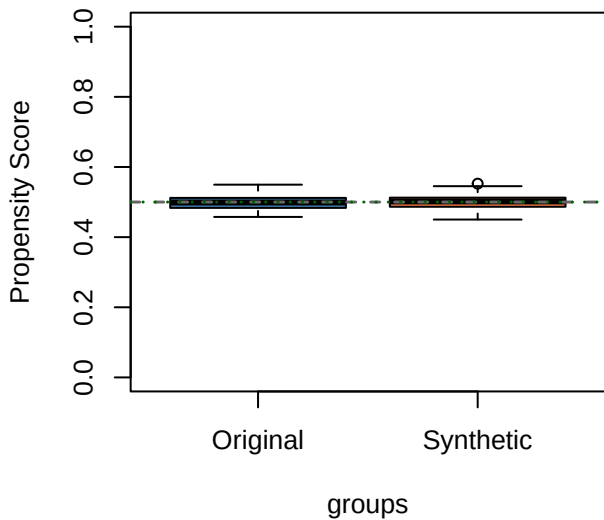
Propensity Score Densities



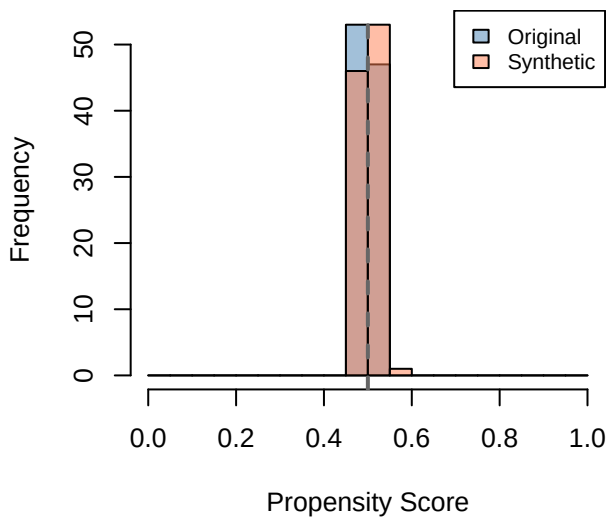
ECDF (SPECKS = 0.1)



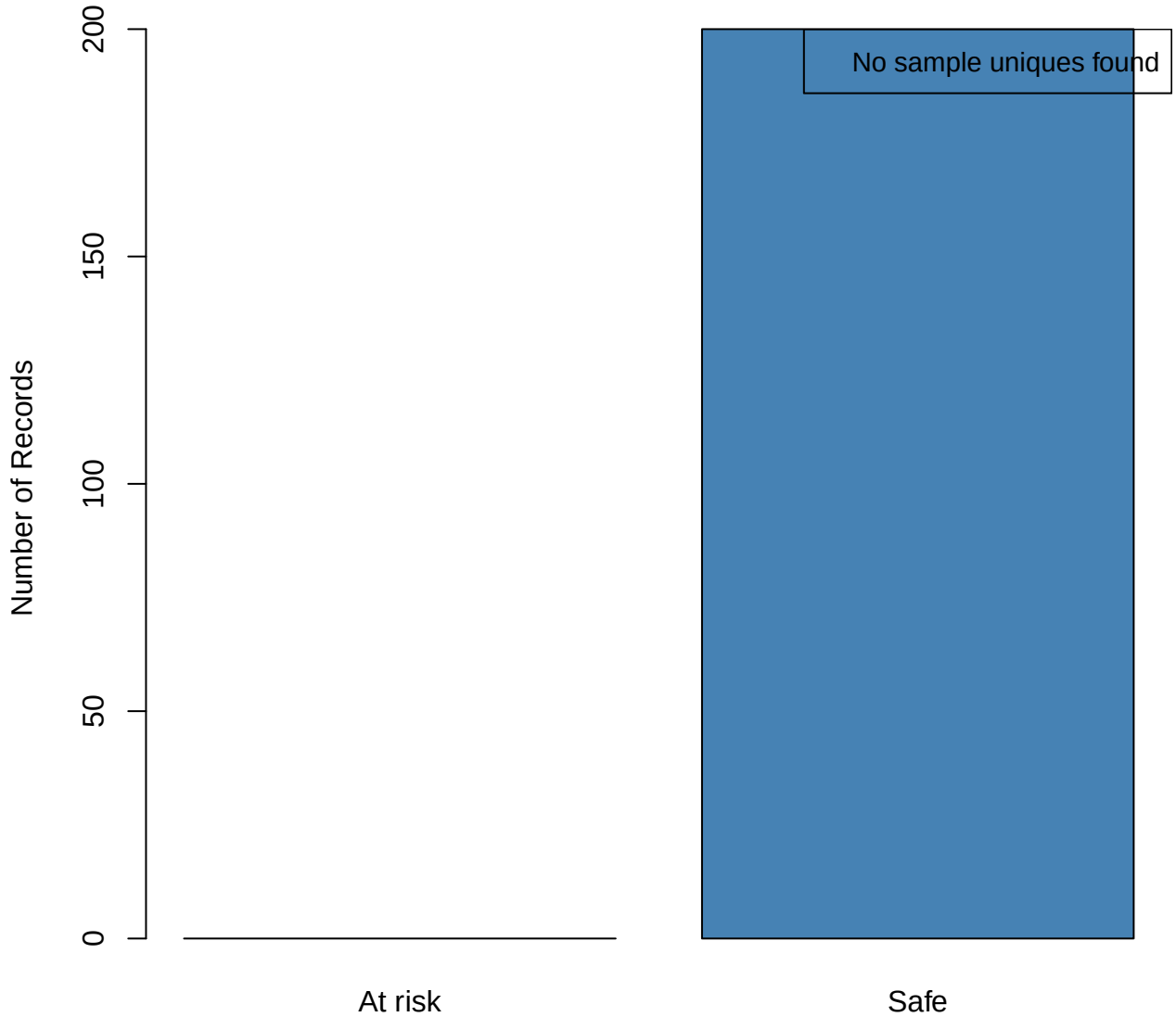
Propensity Scores (PO50 = 53.5%)



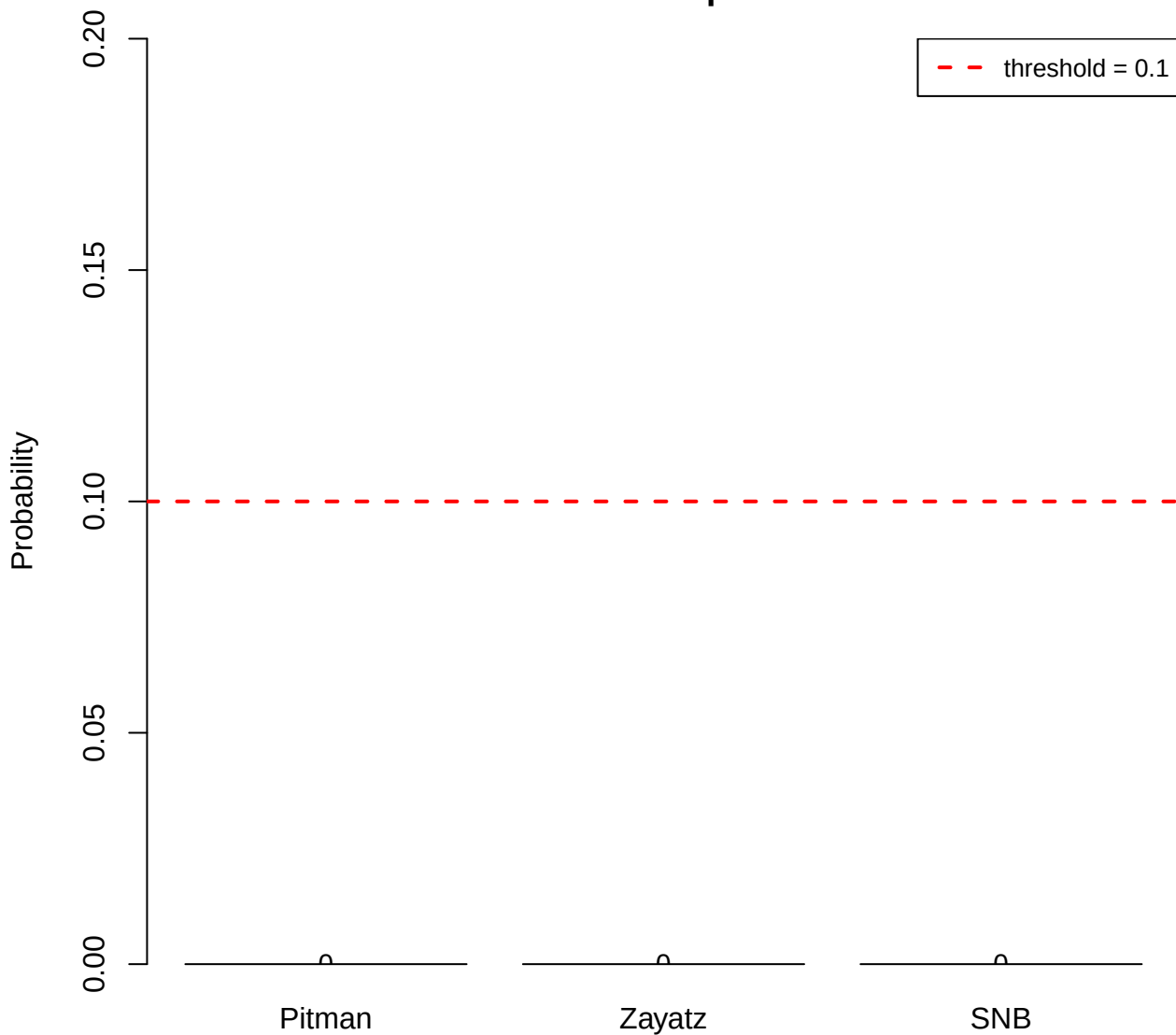
Propensity Score Distributions



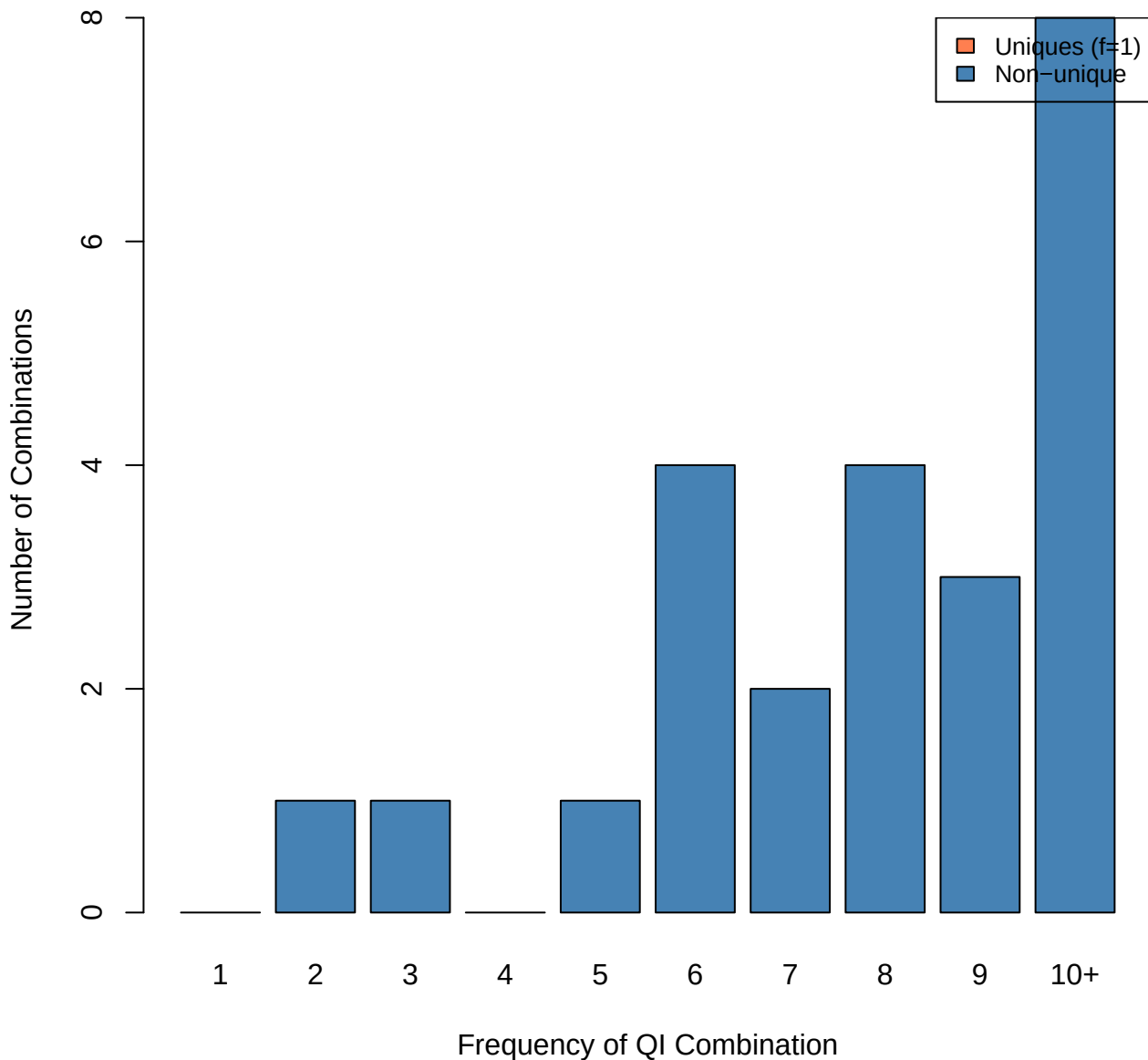
Population Uniqueness Risk (all)



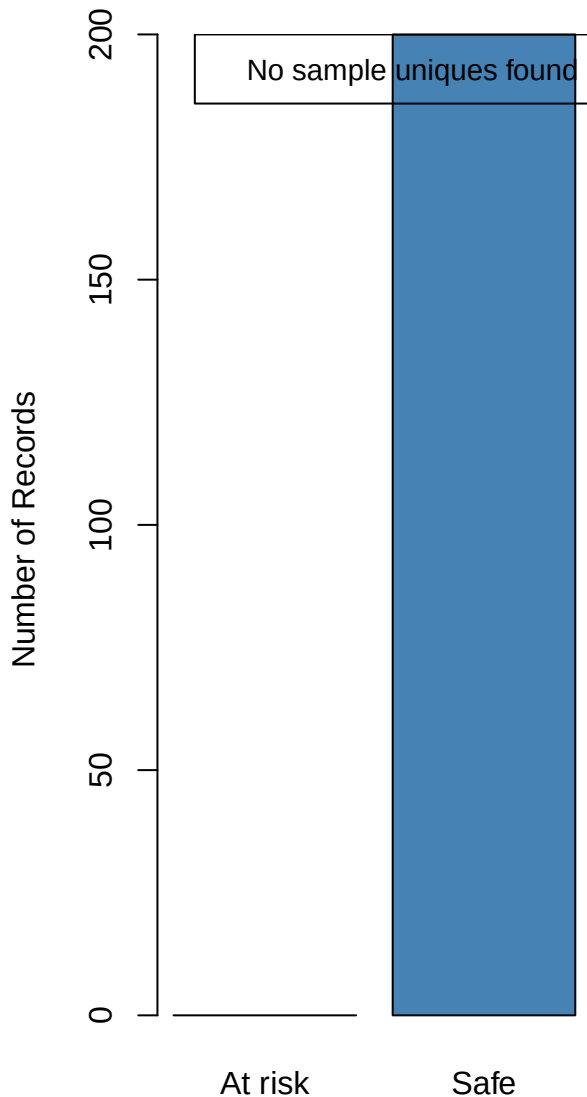
P(pop unique | sample unique) Method Comparison



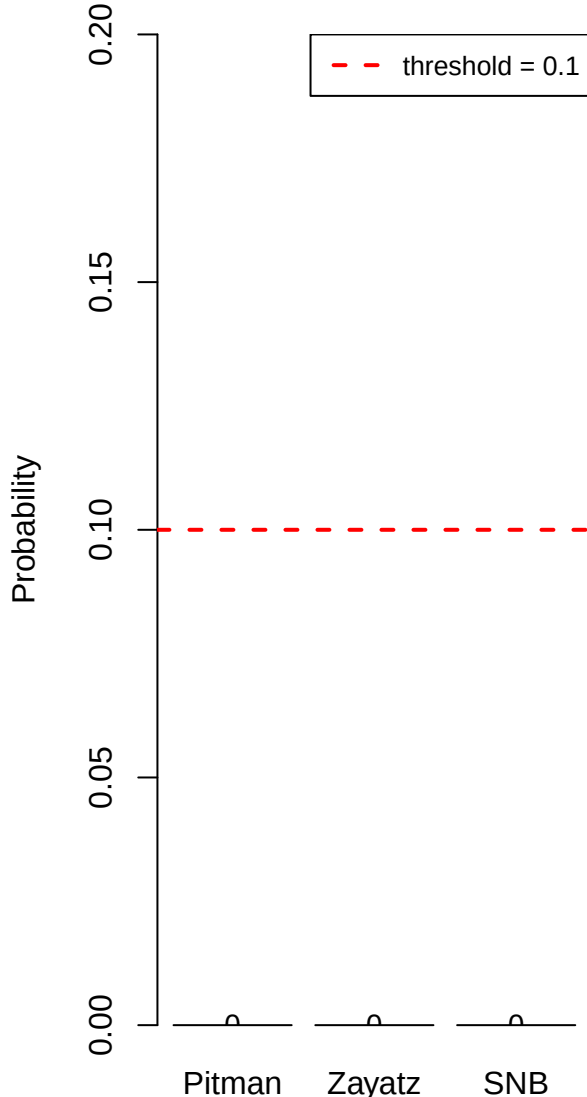
QI Frequency Distribution



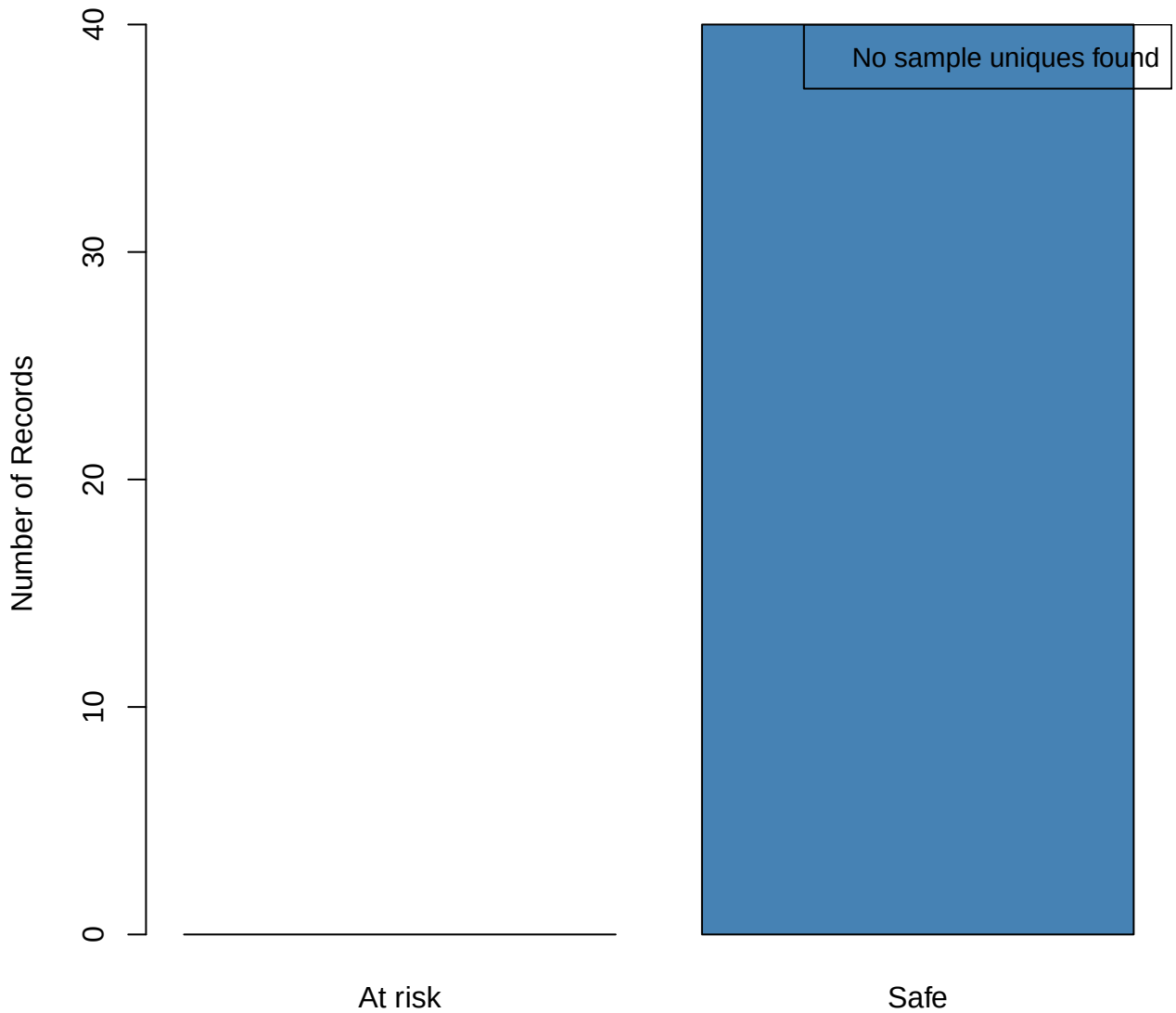
**Population Uniqueness Risk
(all)**



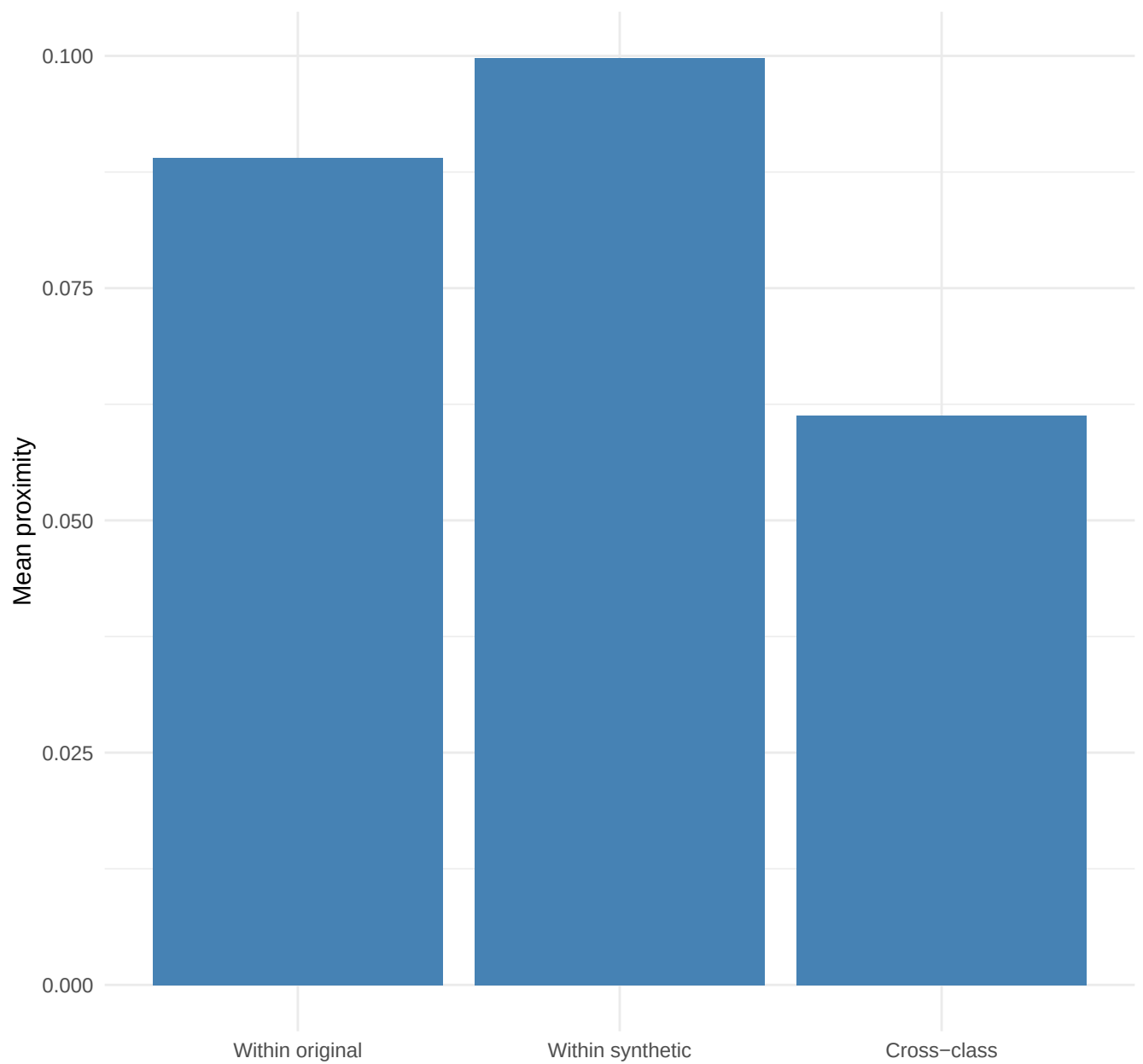
**P(pop unique | sample unique)
Method Comparison**



Population Uniqueness Risk (all)

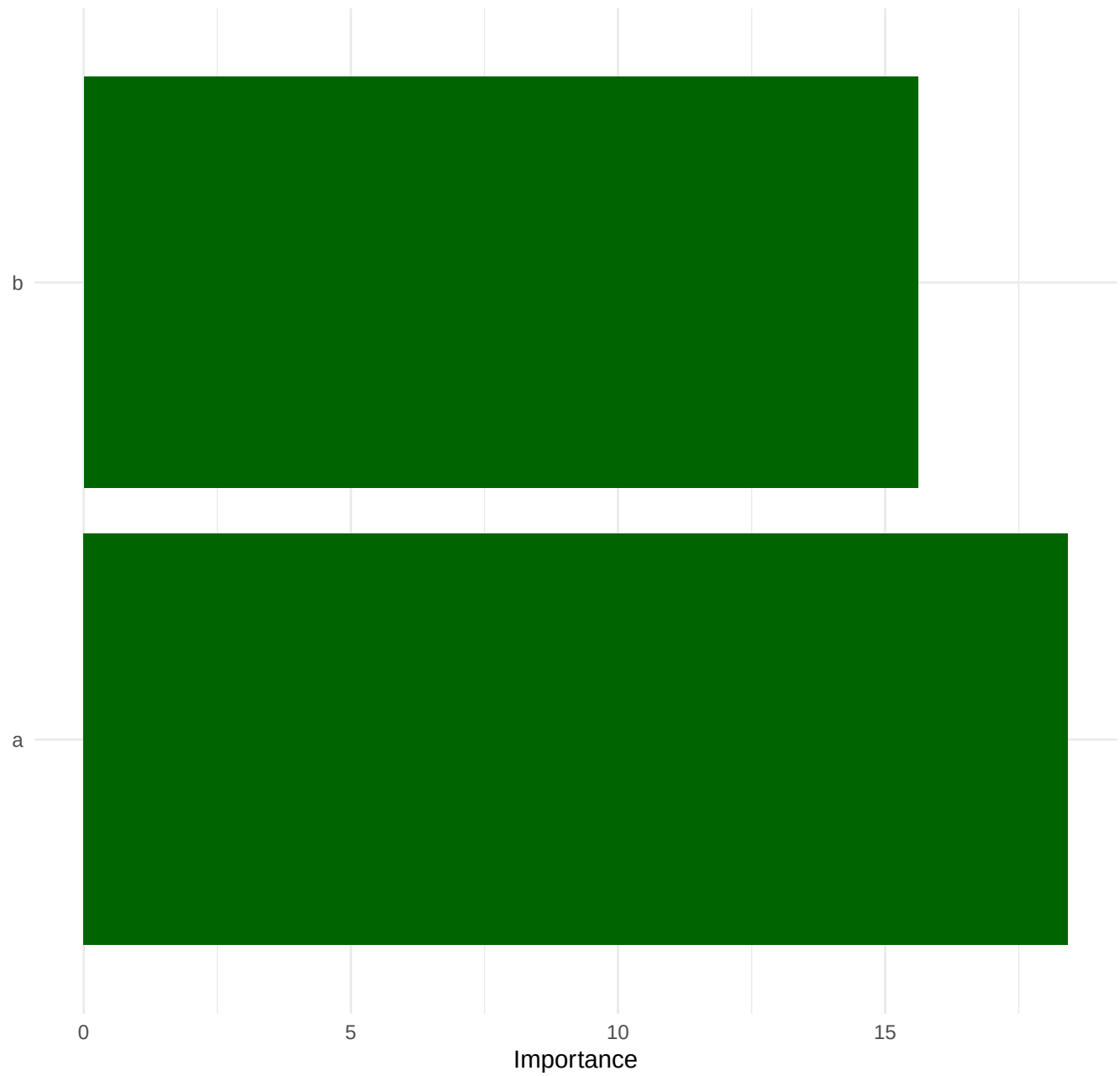


Proximity Structure

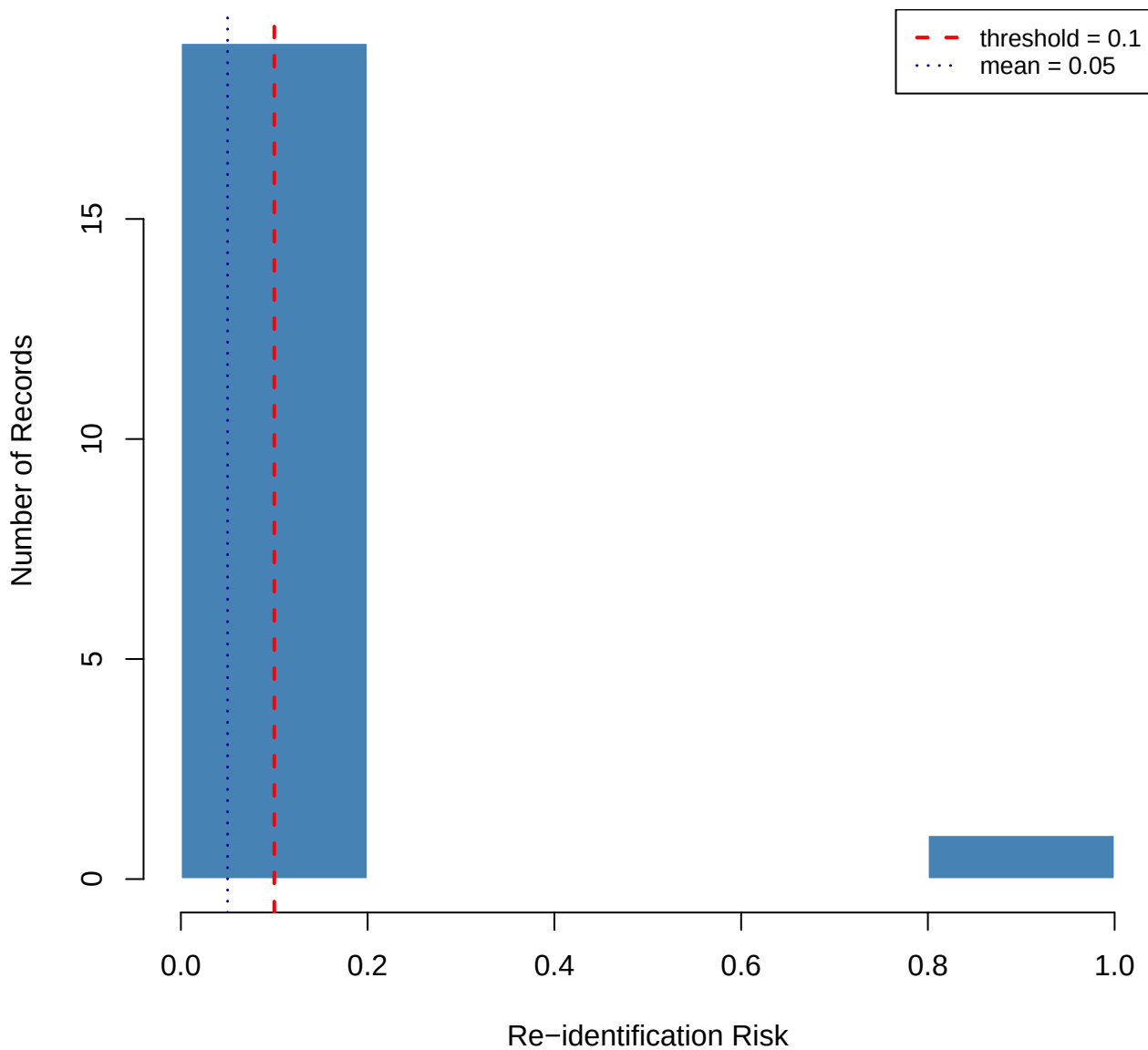


Structure ratio: 0.649

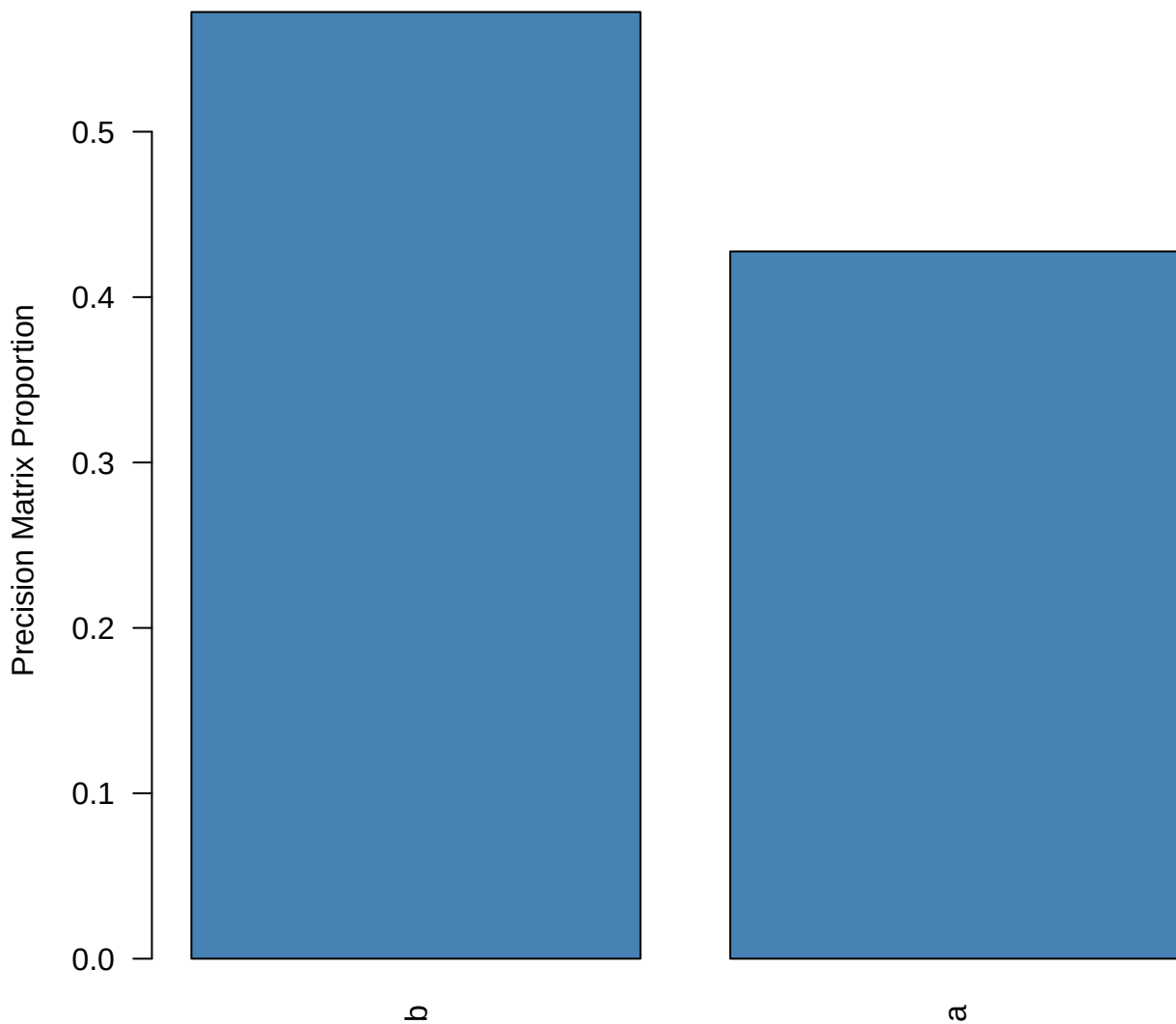
Variable Importance



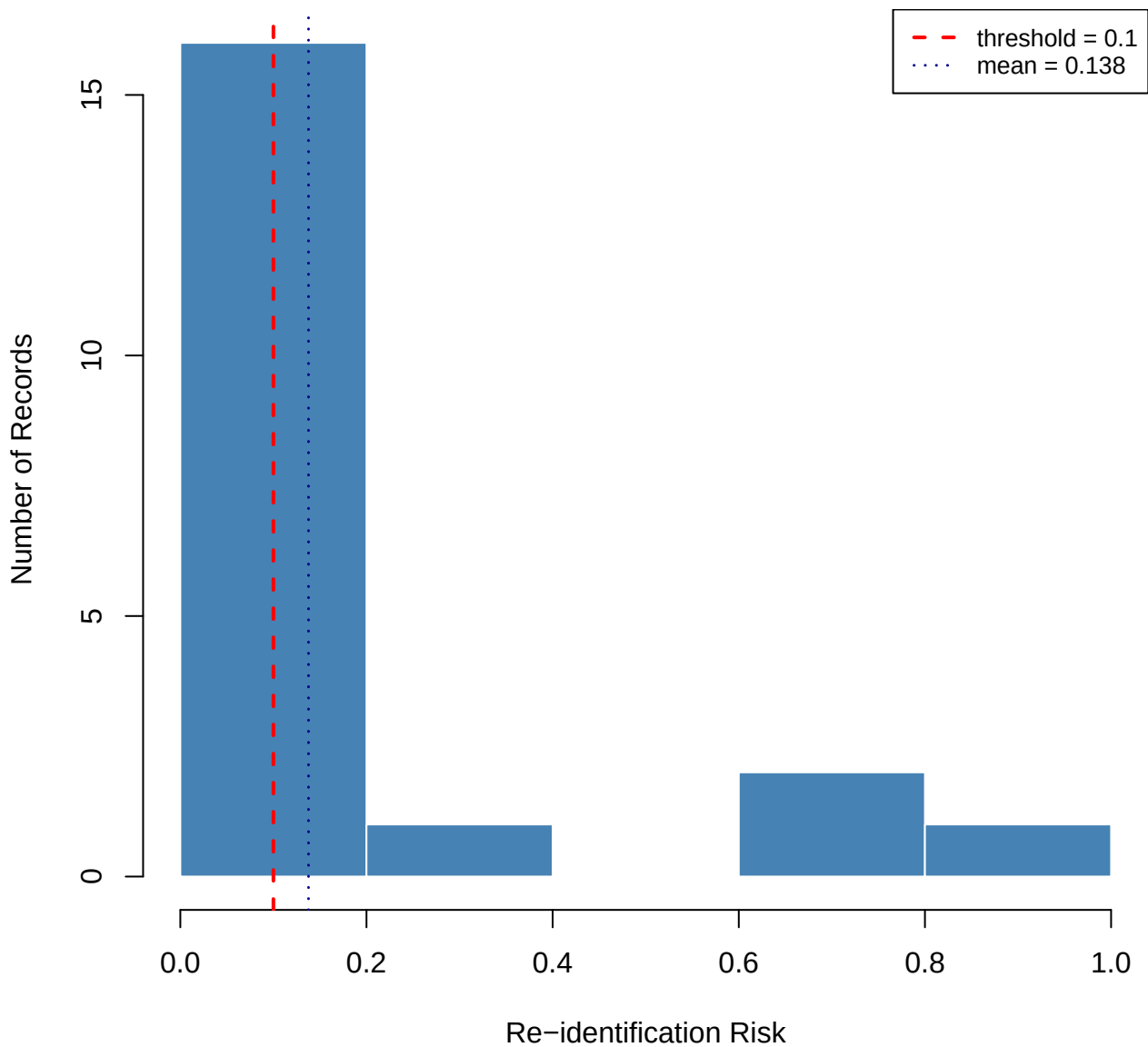
Record Linkage Risk Distribution



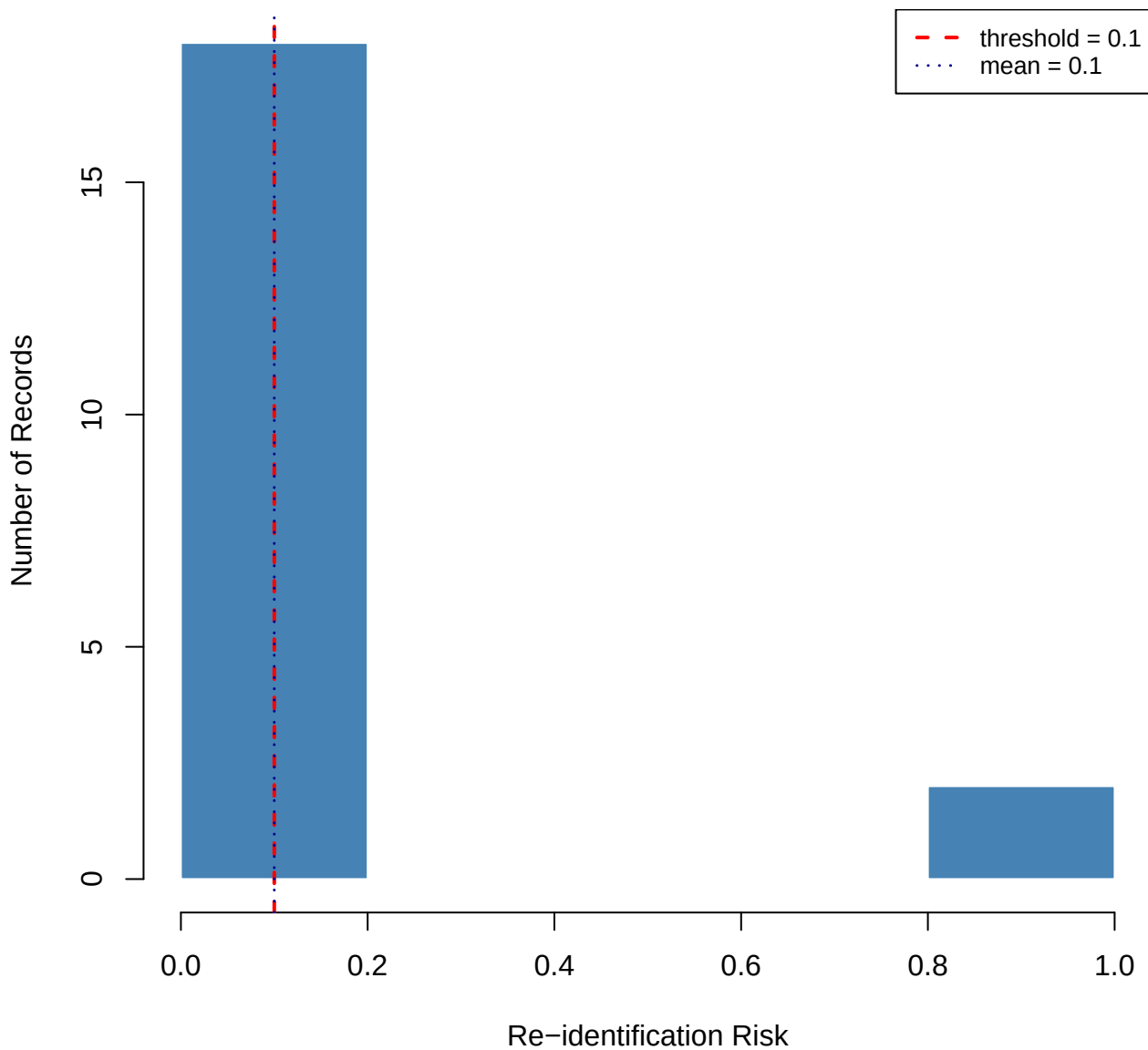
Variable Importance



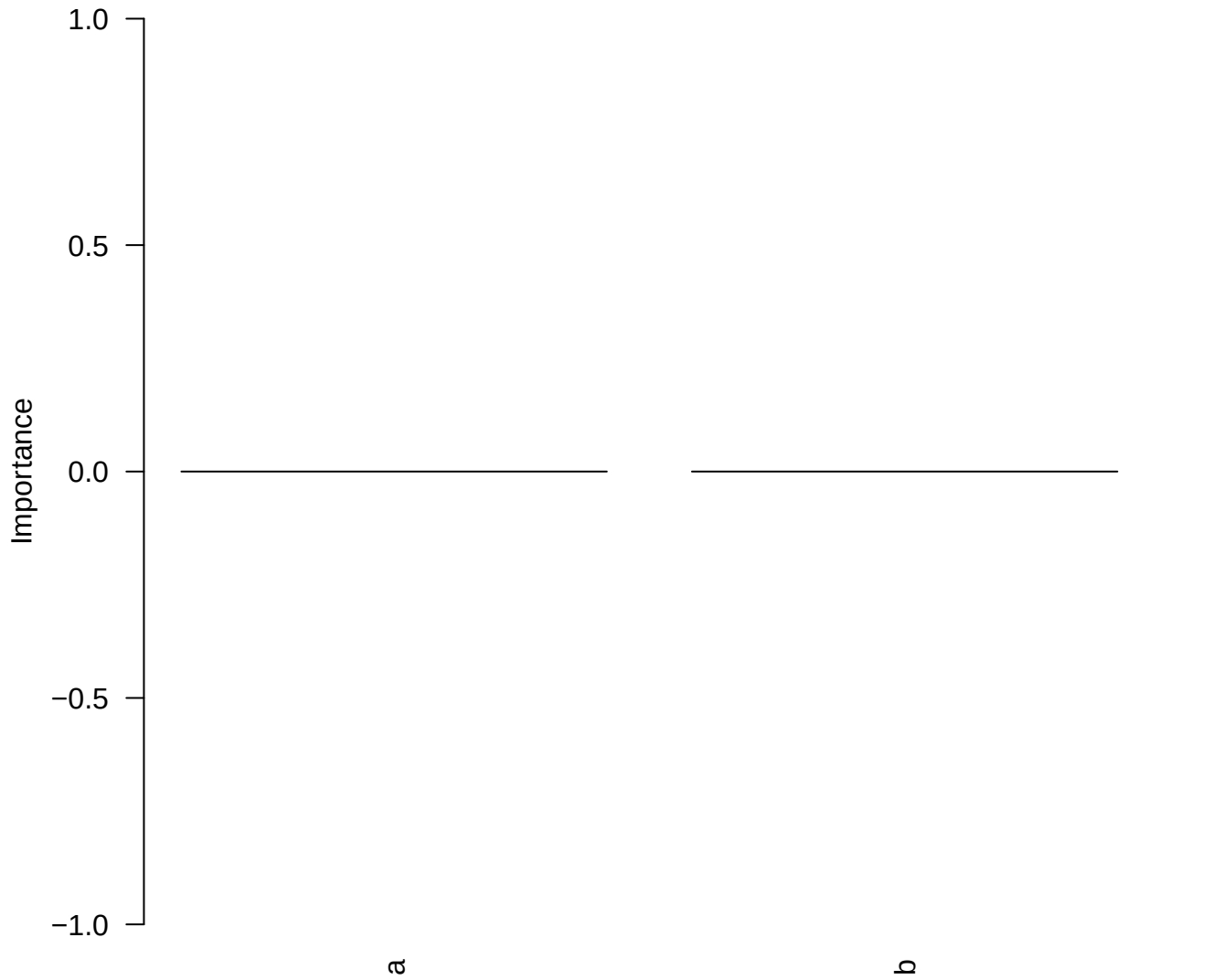
Record Linkage Risk Distribution



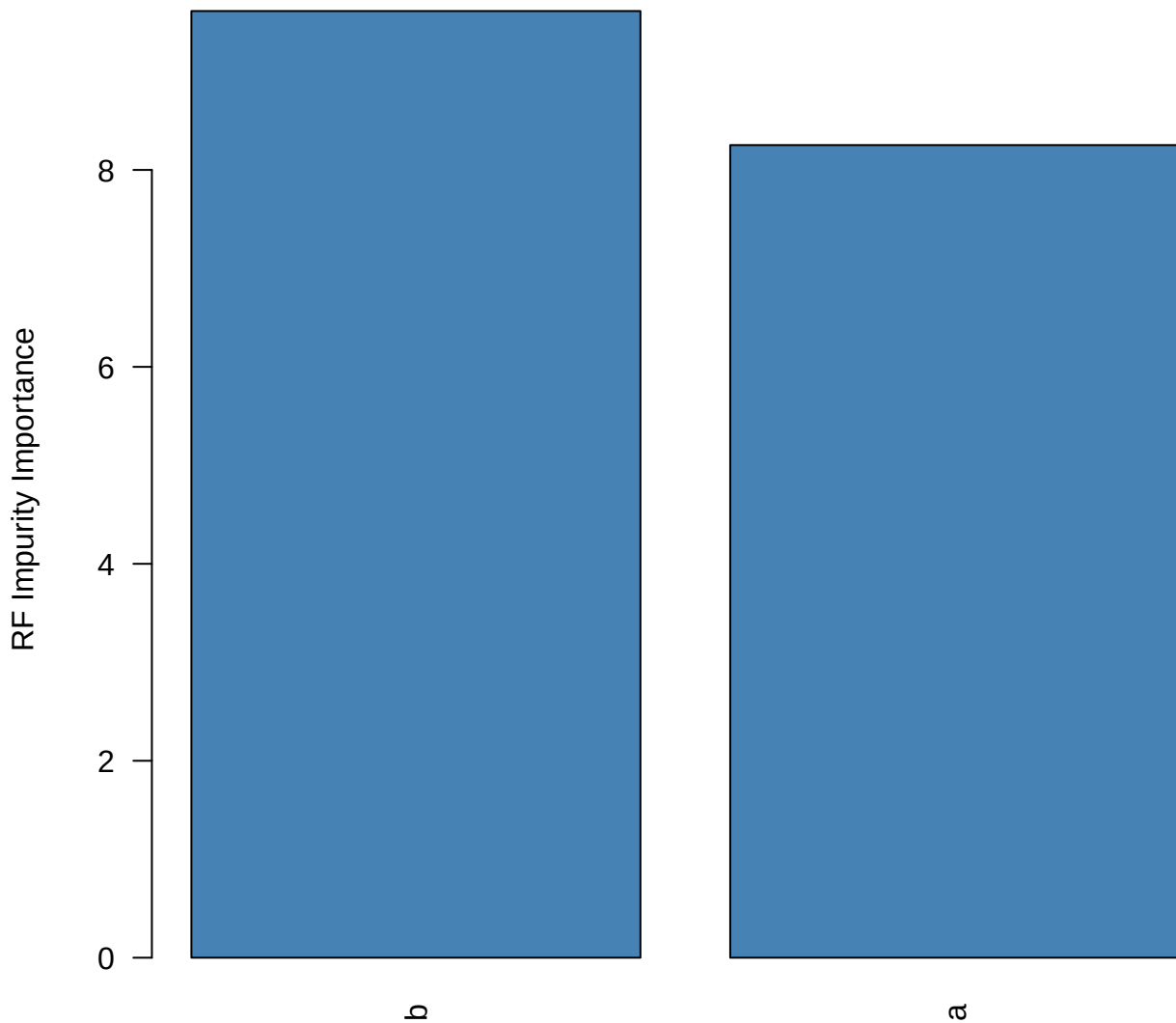
Record Linkage Risk Distribution



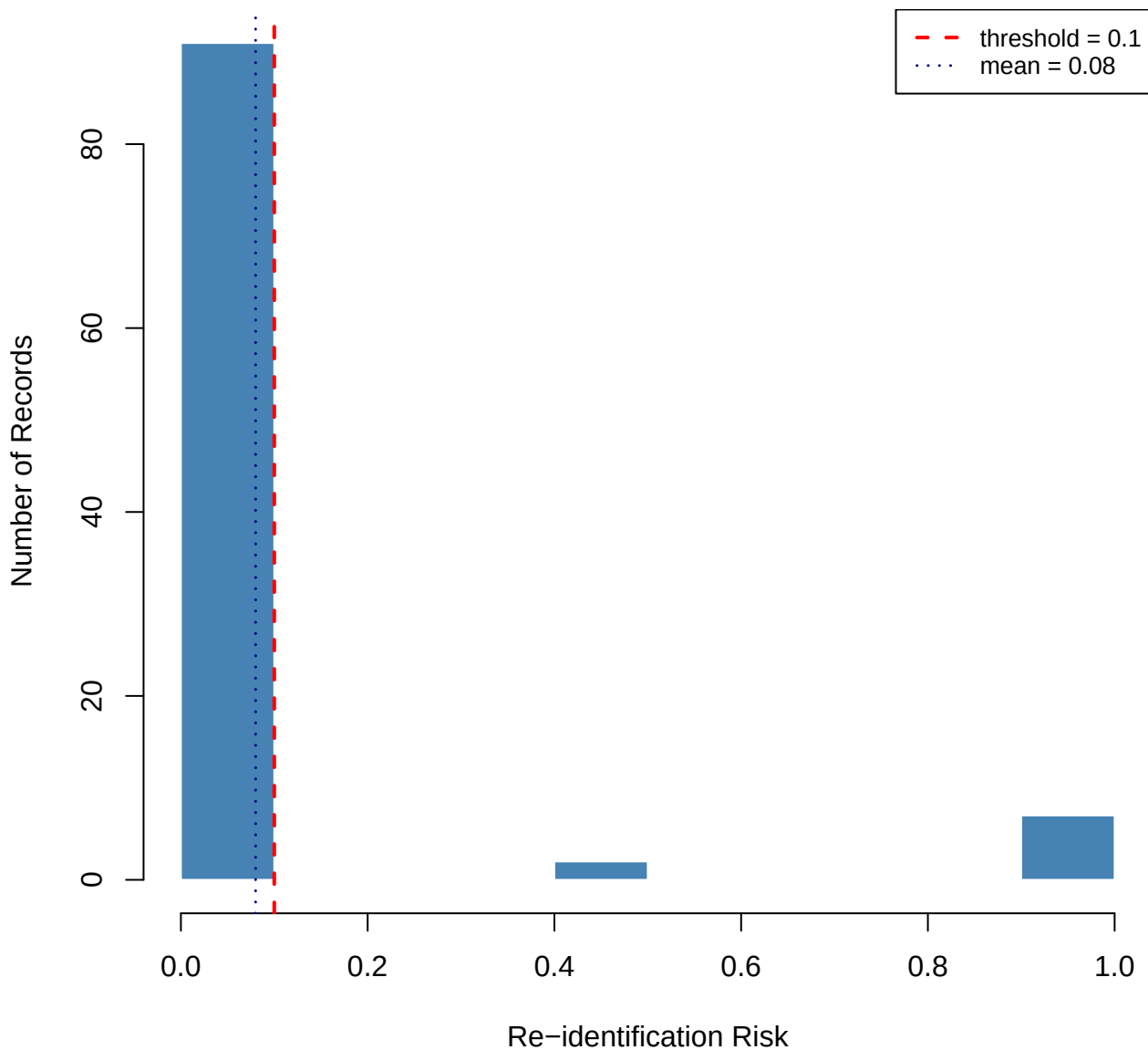
Variable Importance



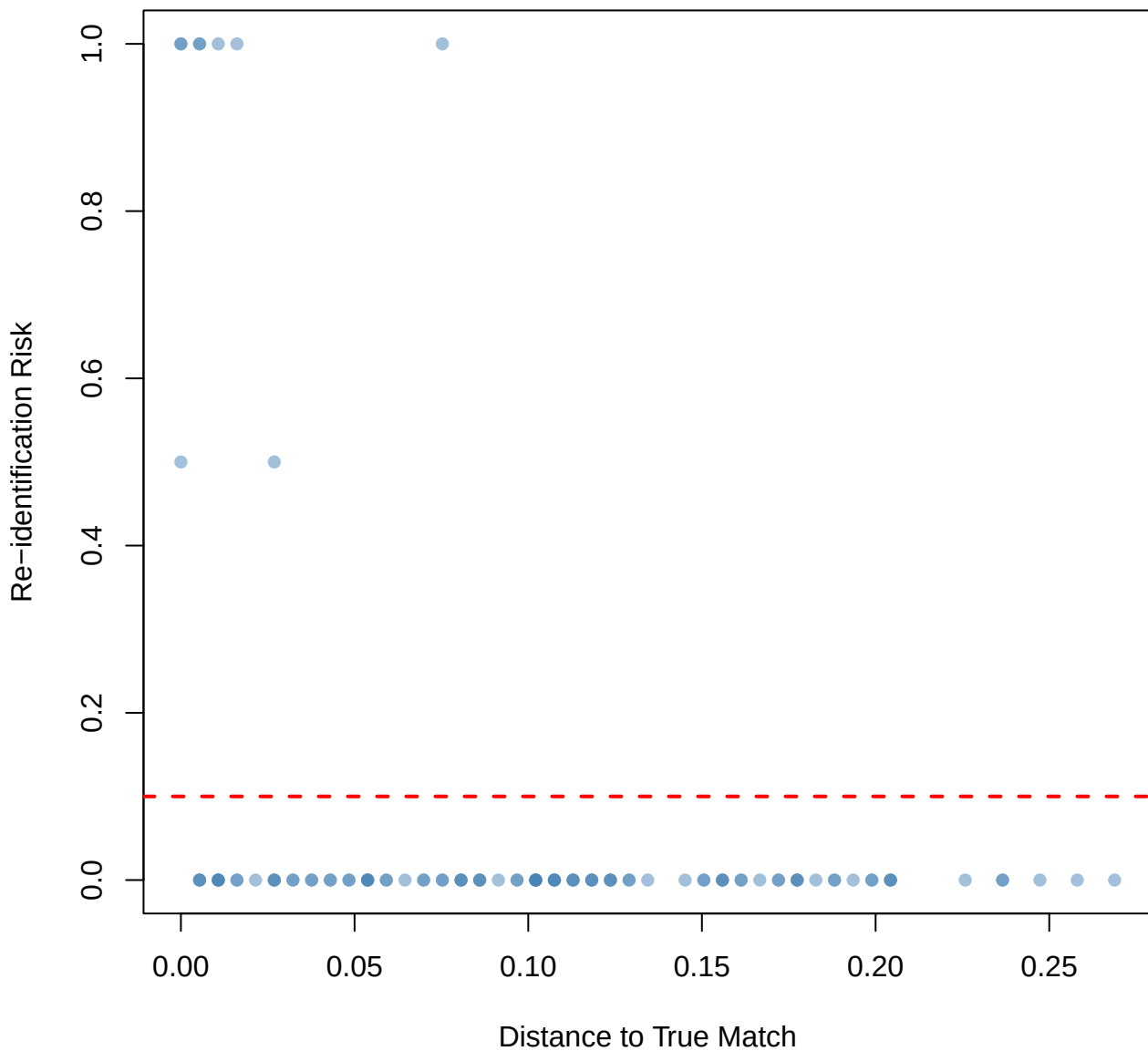
Variable Importance



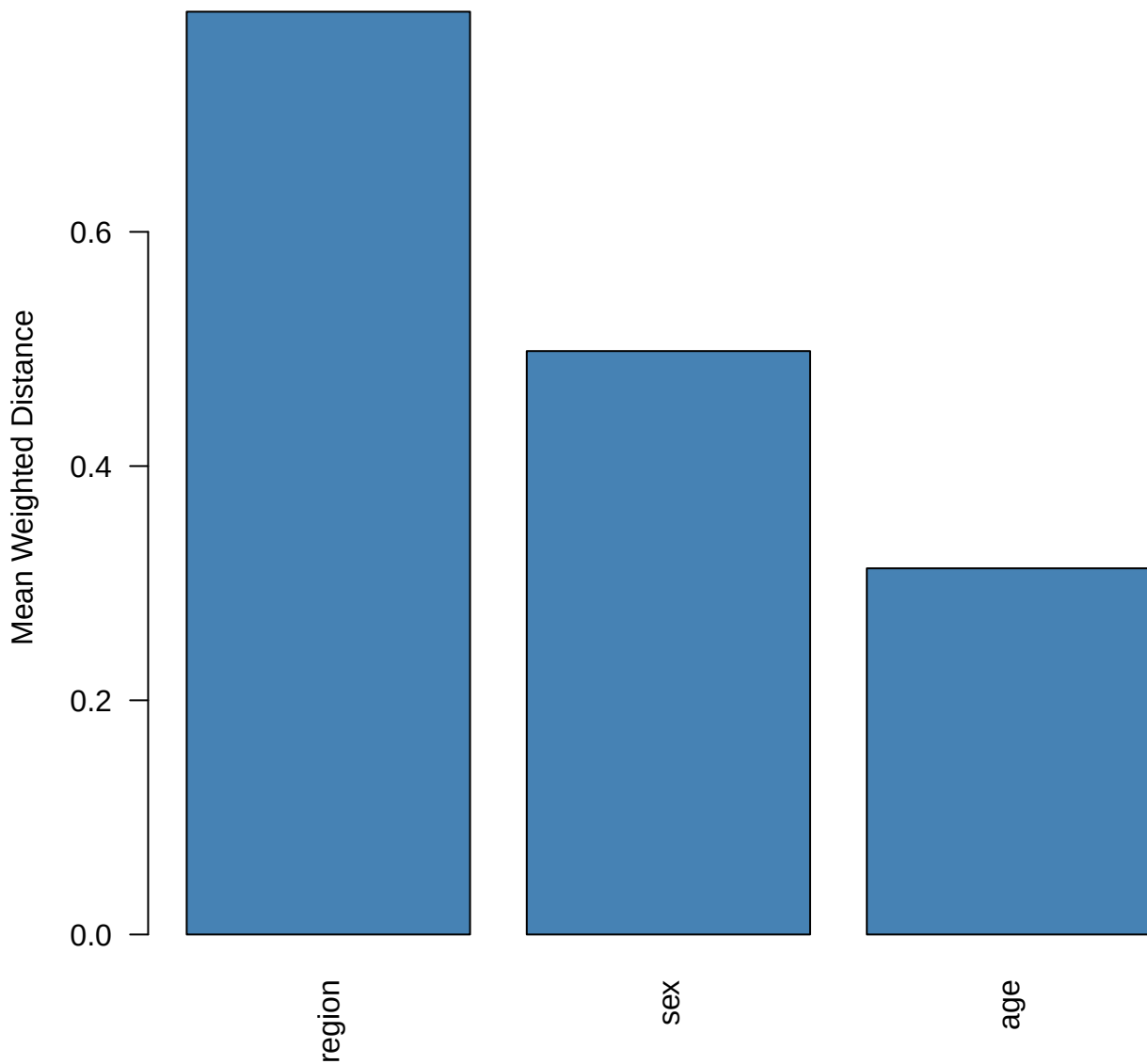
Record Linkage Risk Distribution



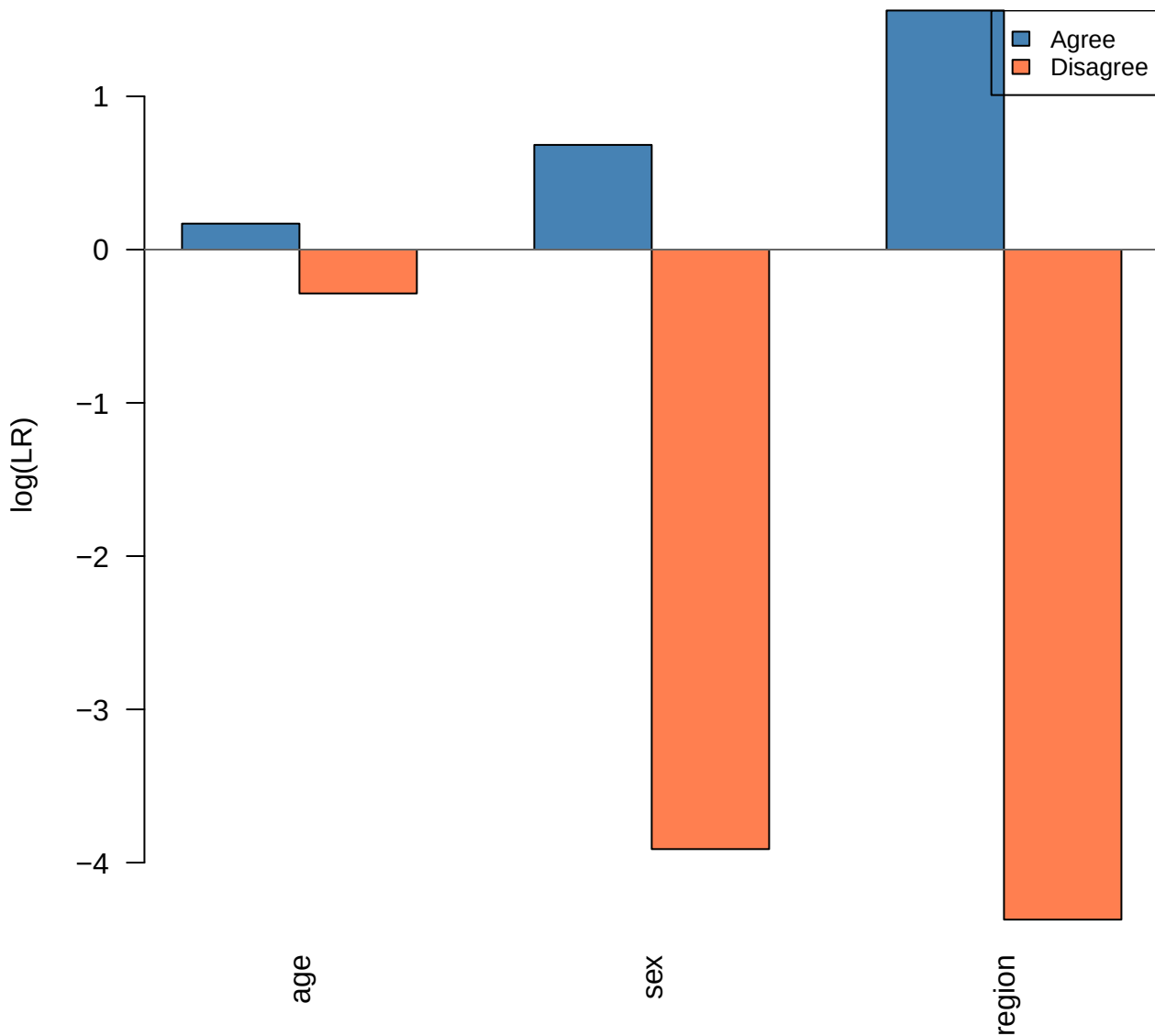
Distance to True Match vs Risk



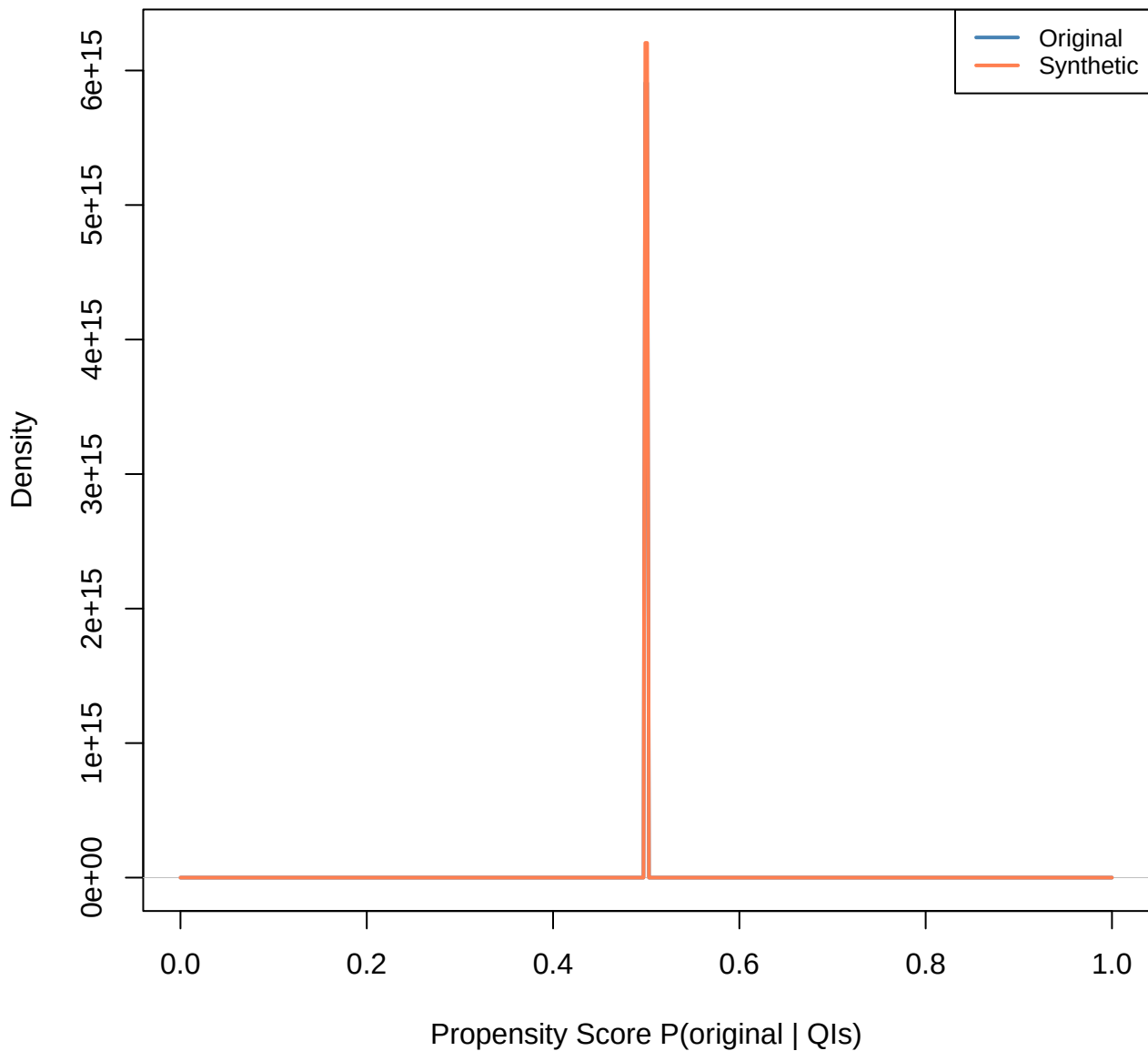
Variable Importance



Per-Variable Log-Likelihood Ratios

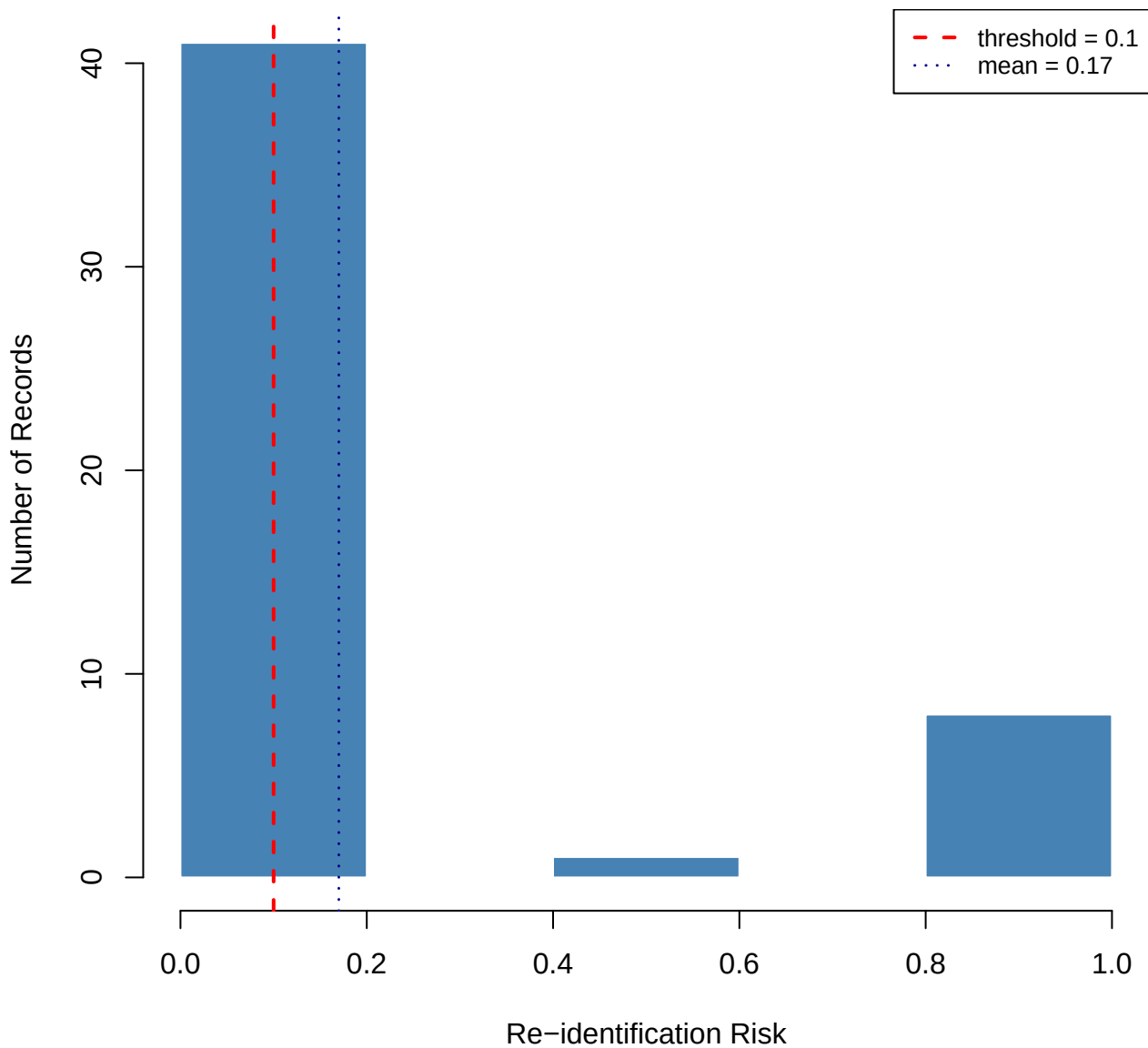


Propensity Score Distributions

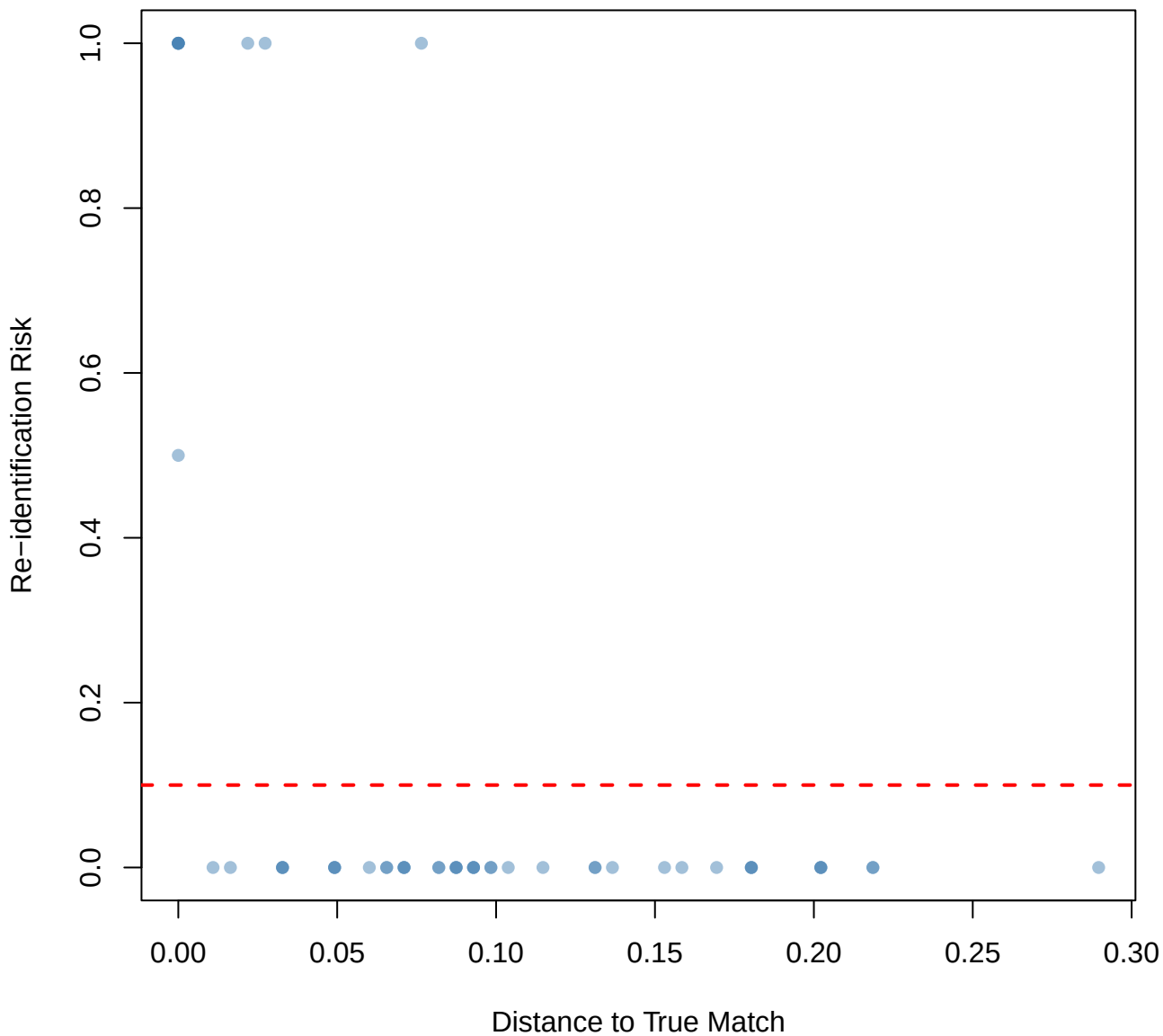


Plot 4 requires method='predictive'

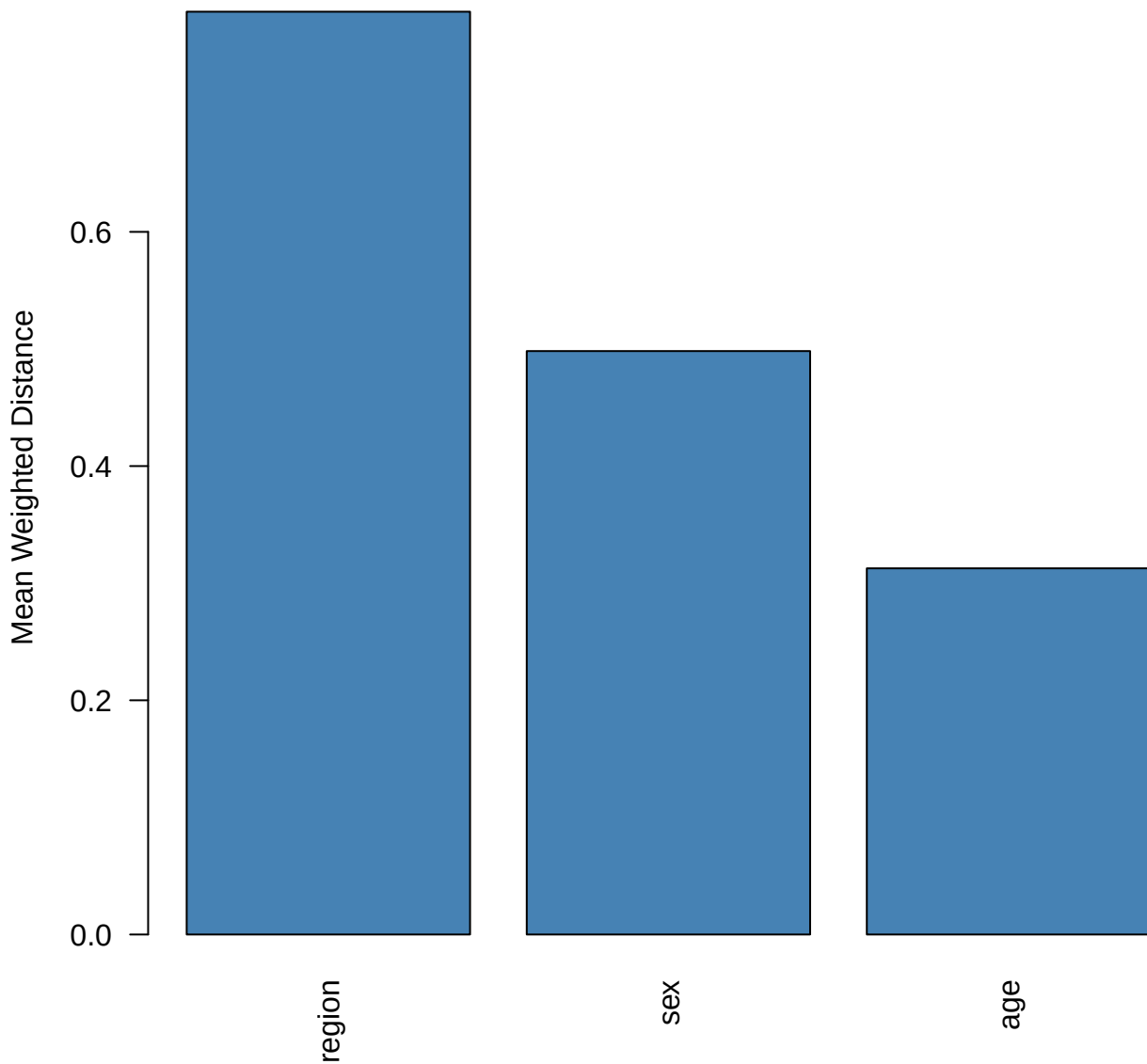
Record Linkage Risk Distribution (reverse)



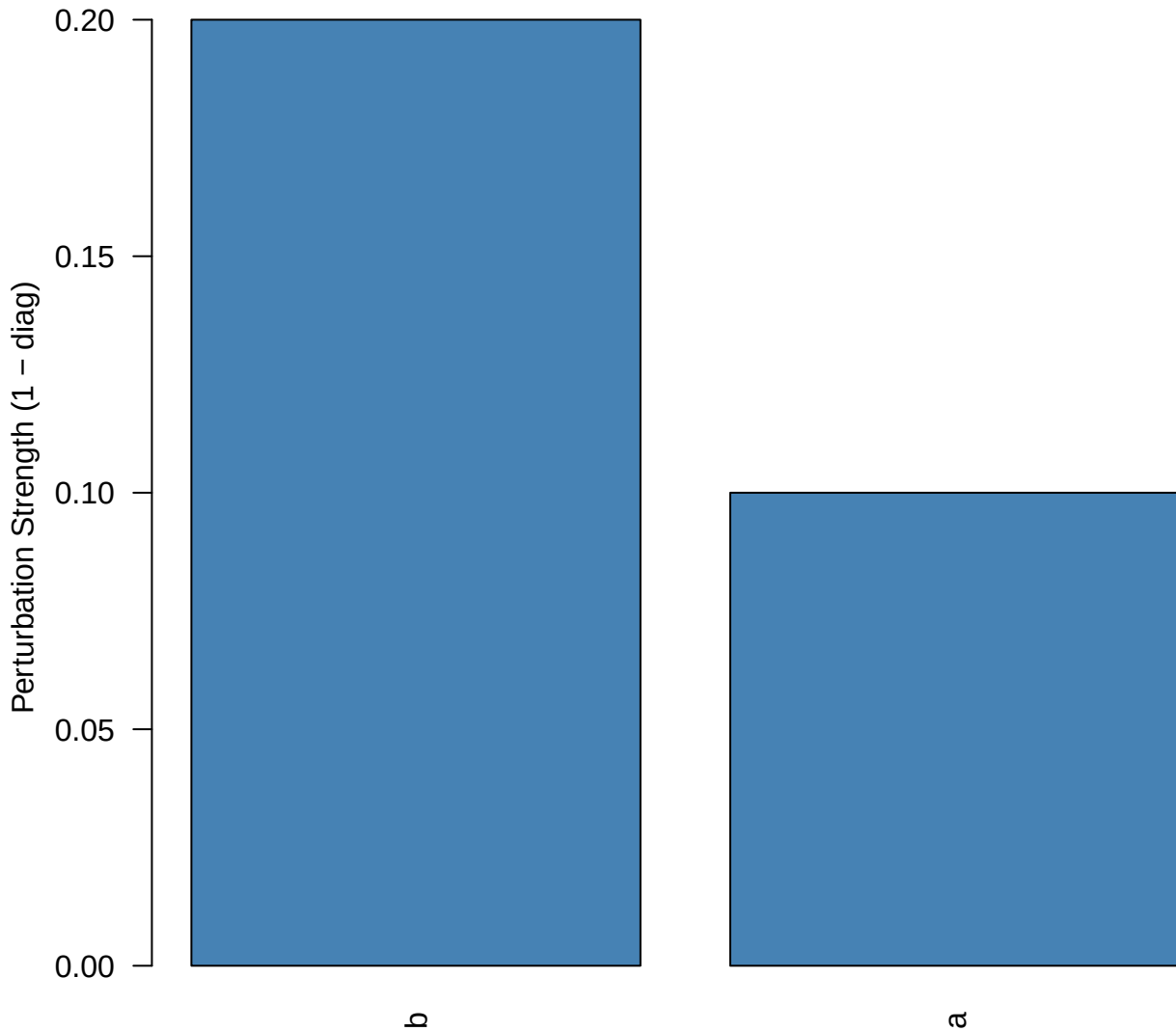
Distance to True Match vs Risk (reverse)



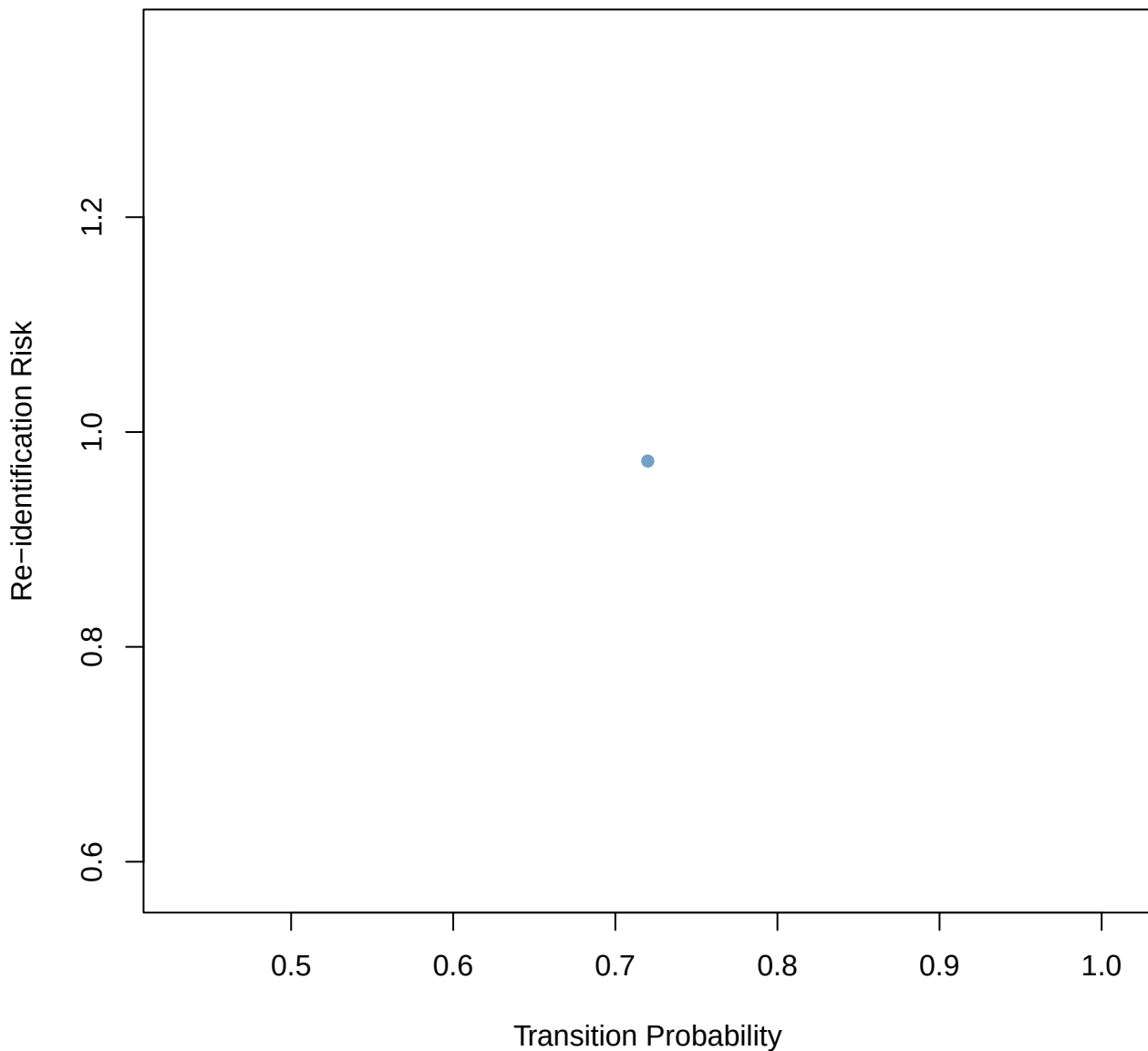
Variable Importance



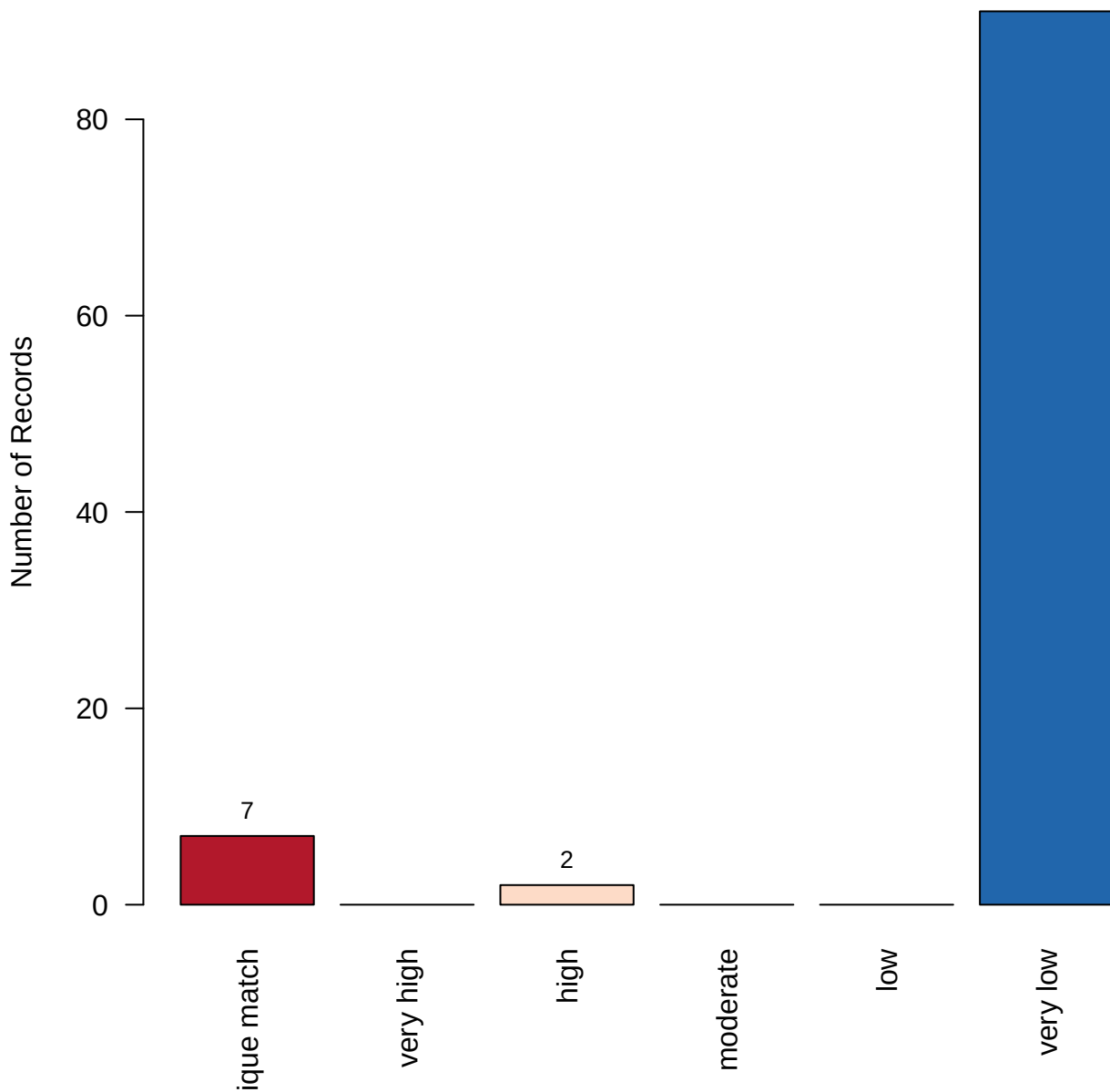
Variable Importance



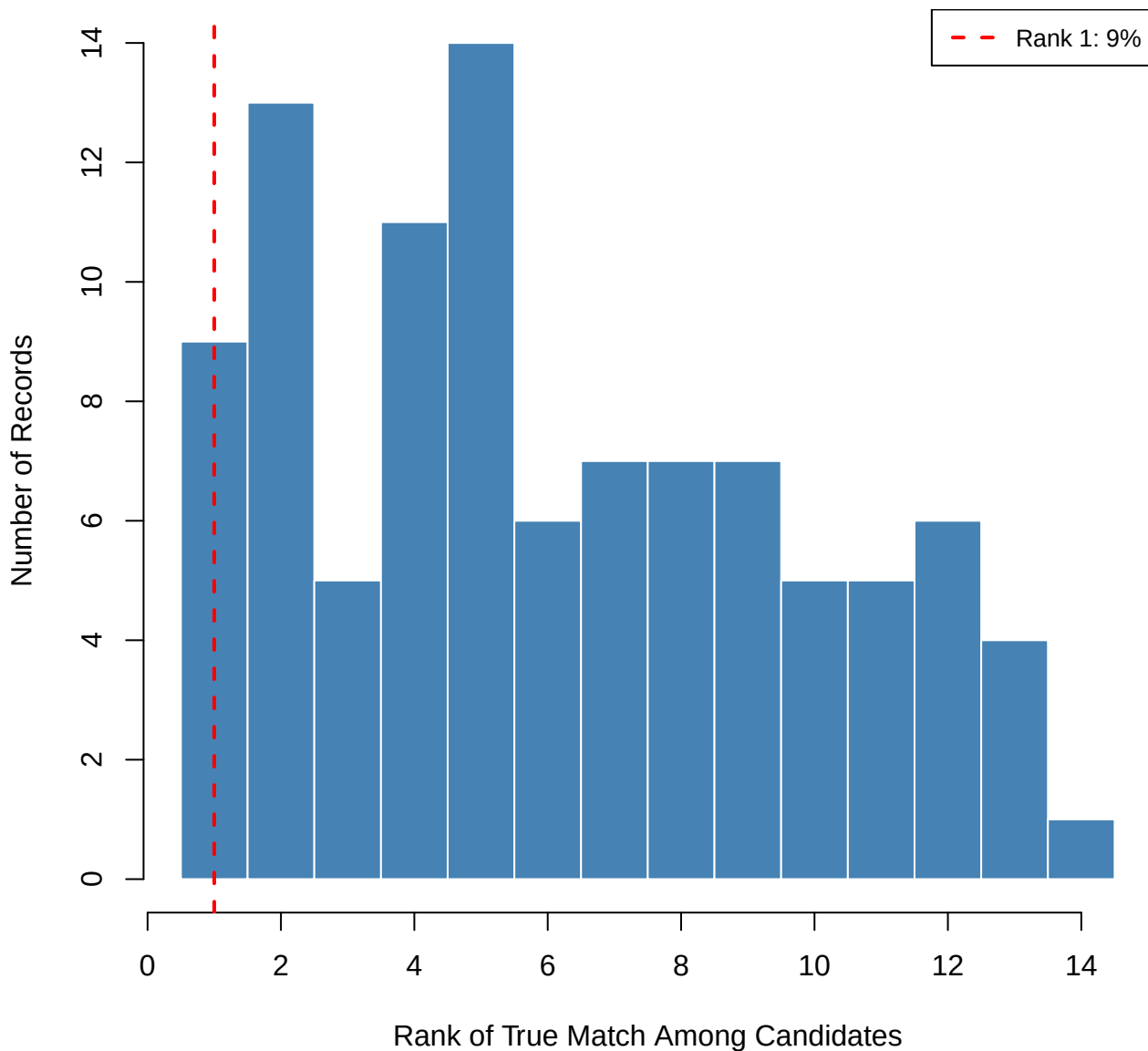
Transition Probability vs Risk



Risk Band Distribution

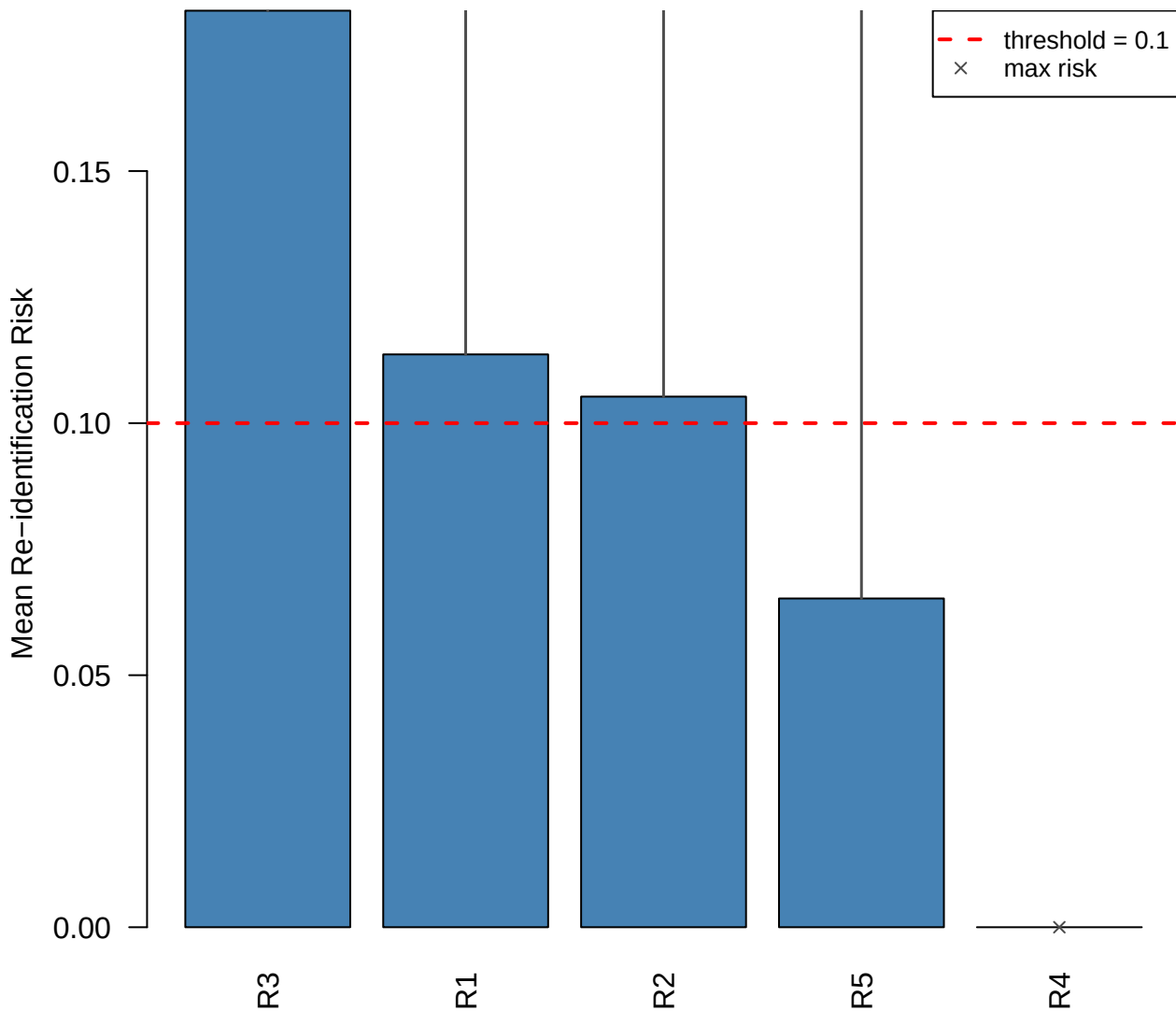


True-Match Rank Distribution

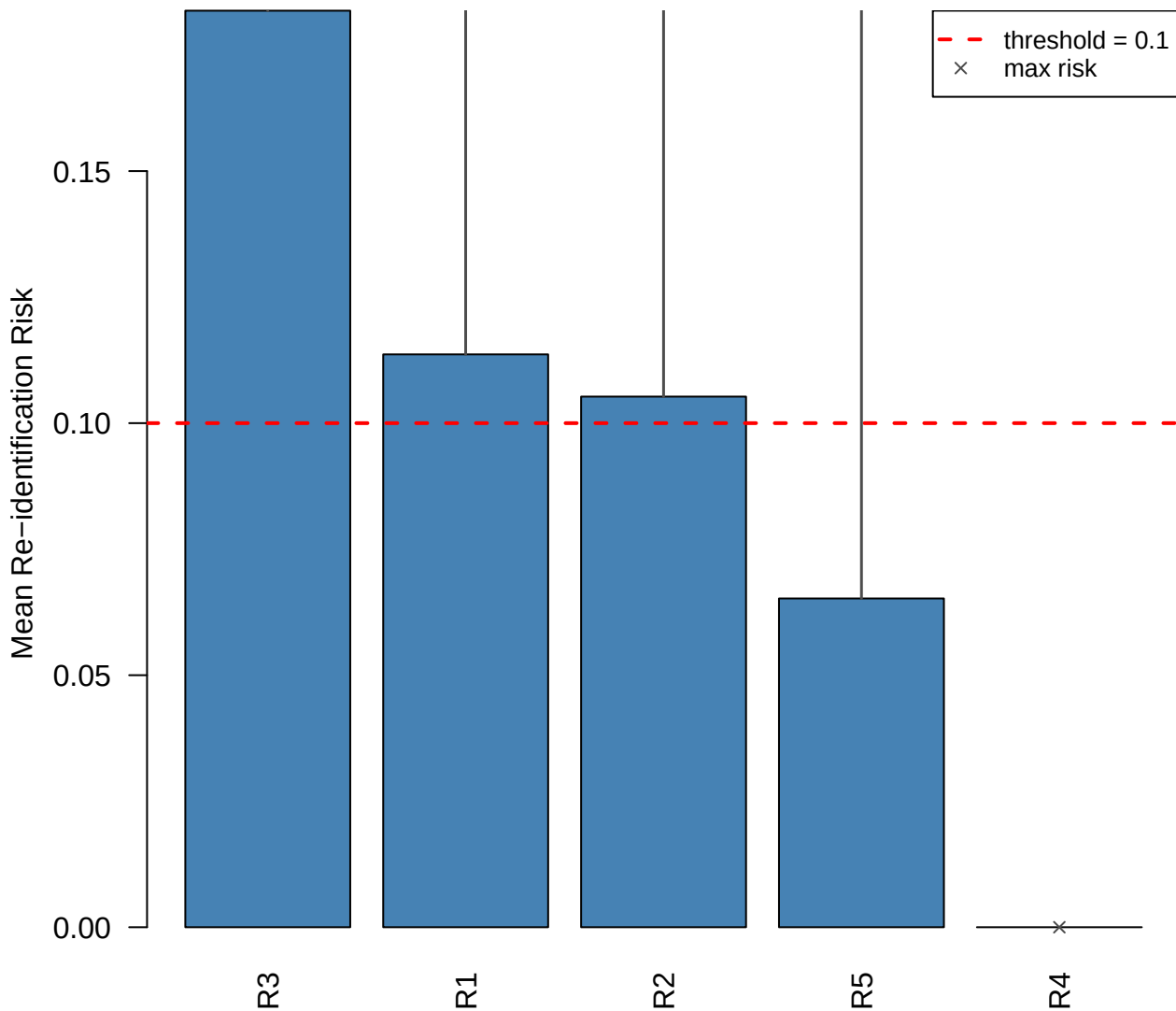


No rank data available
(no true matches found)

Mean Risk by Group

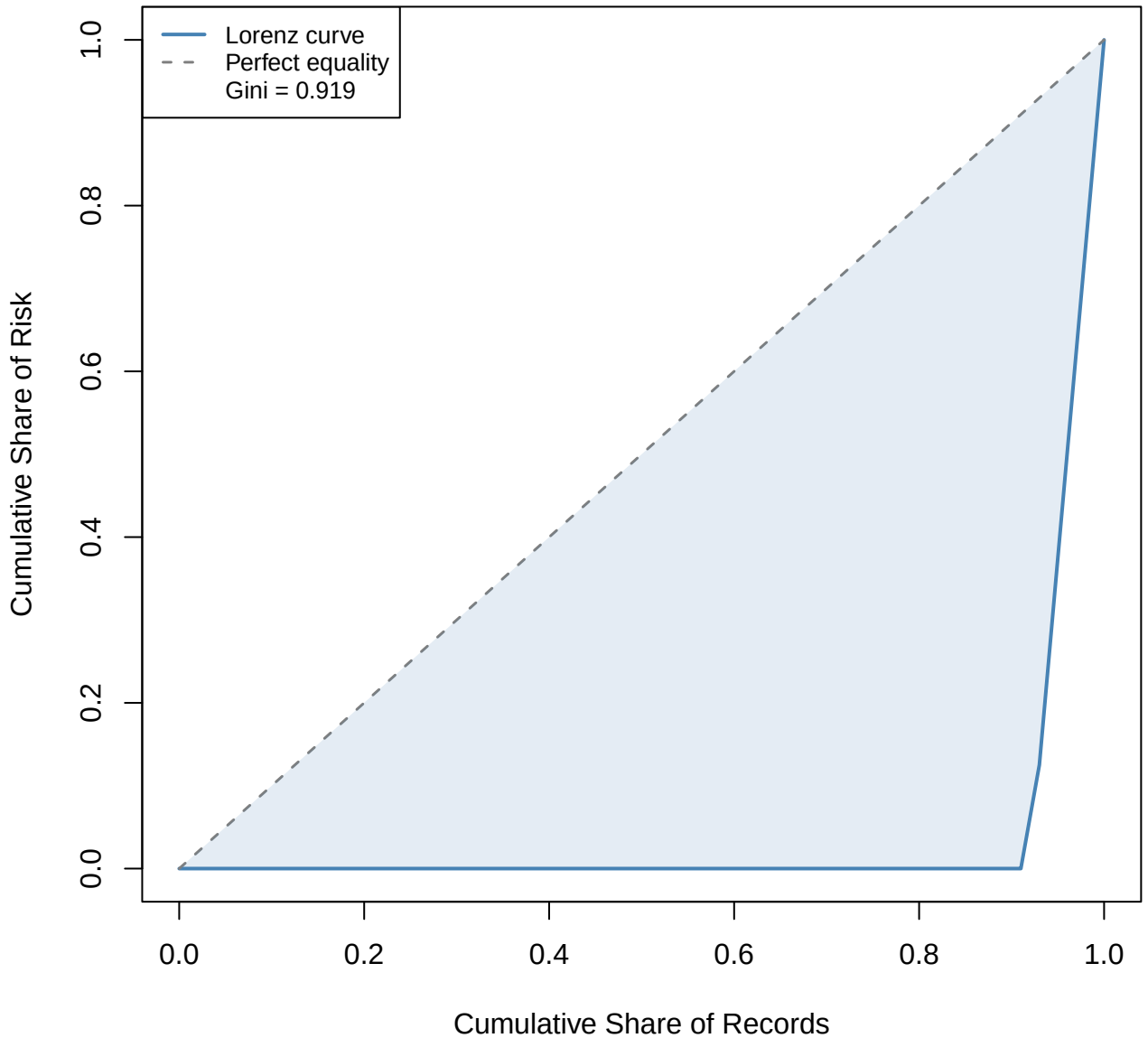


Mean Risk by Group

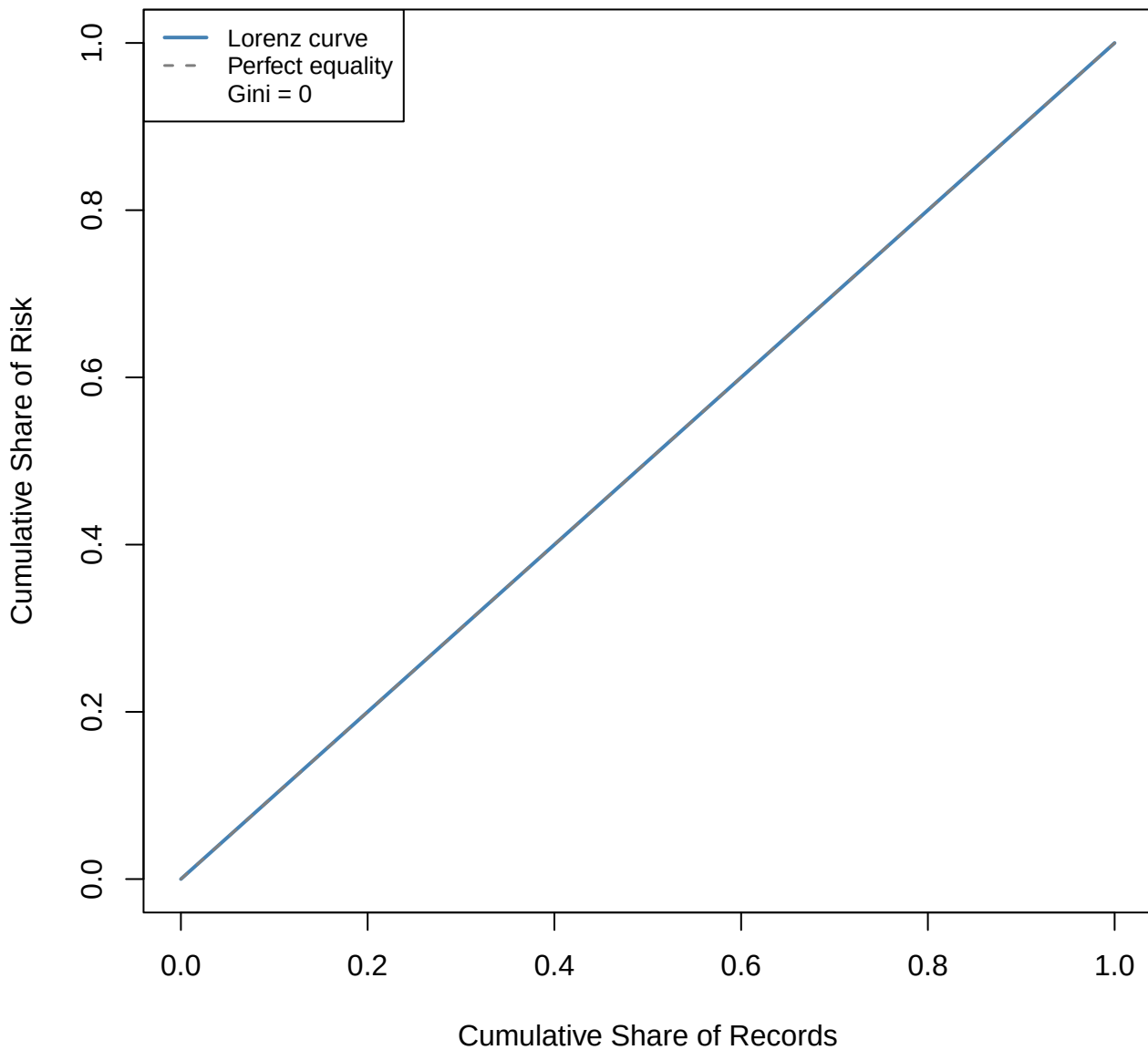


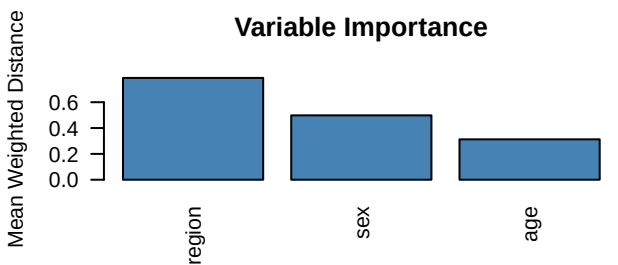
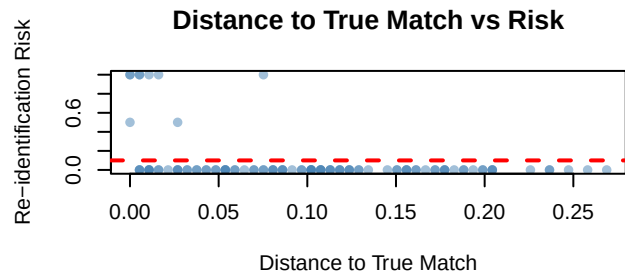
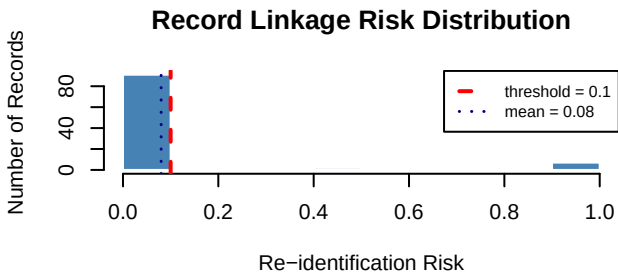
Plot 7 requires 'group' argument

Lorenz Curve of Risk

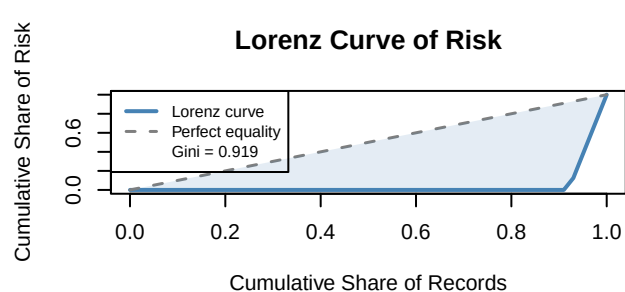
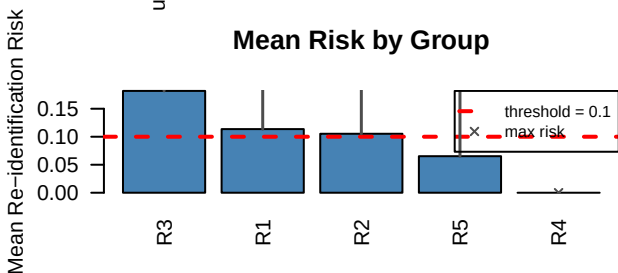
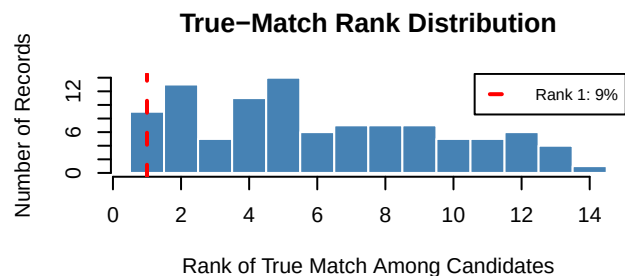
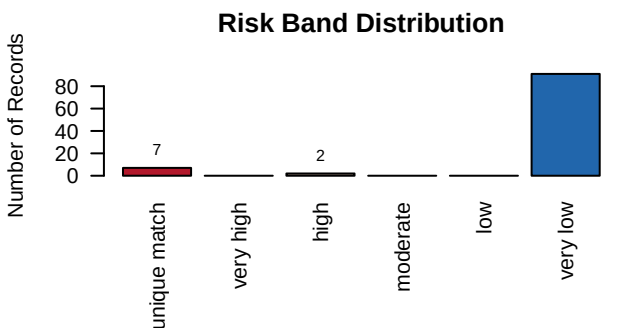


Lorenz Curve of Risk

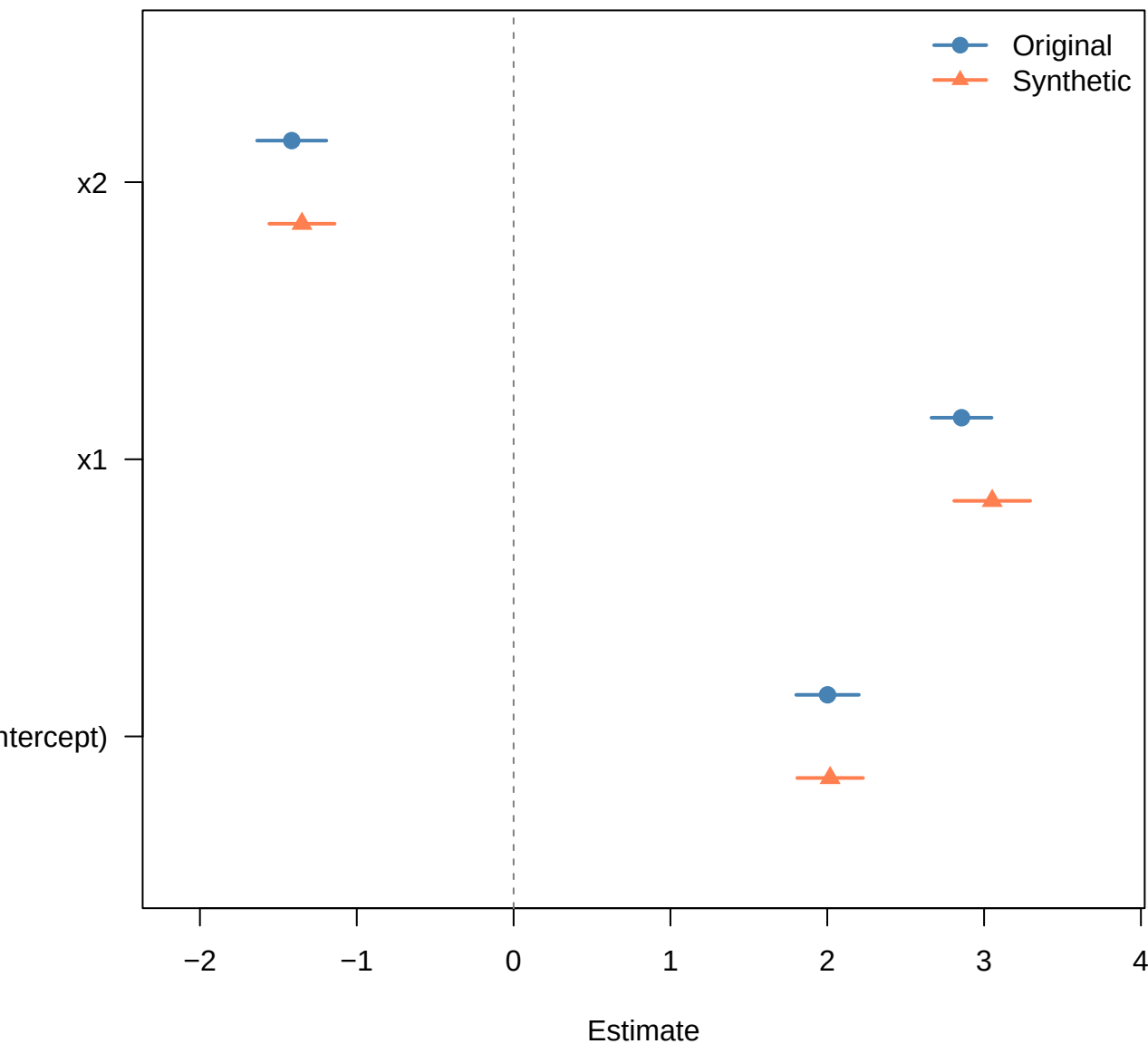




Plot 4 requires method='predictive'

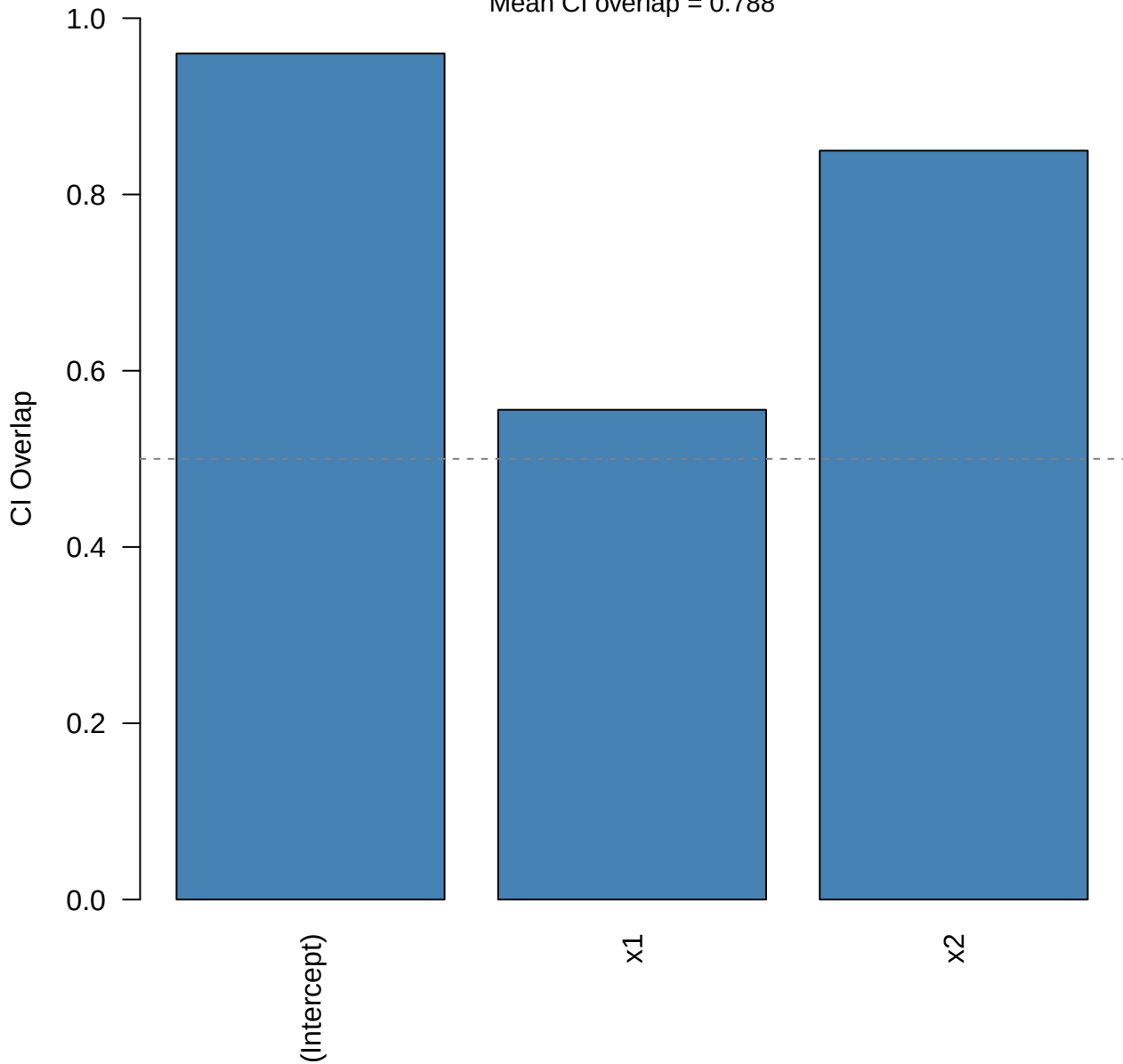


Regression Coefficients

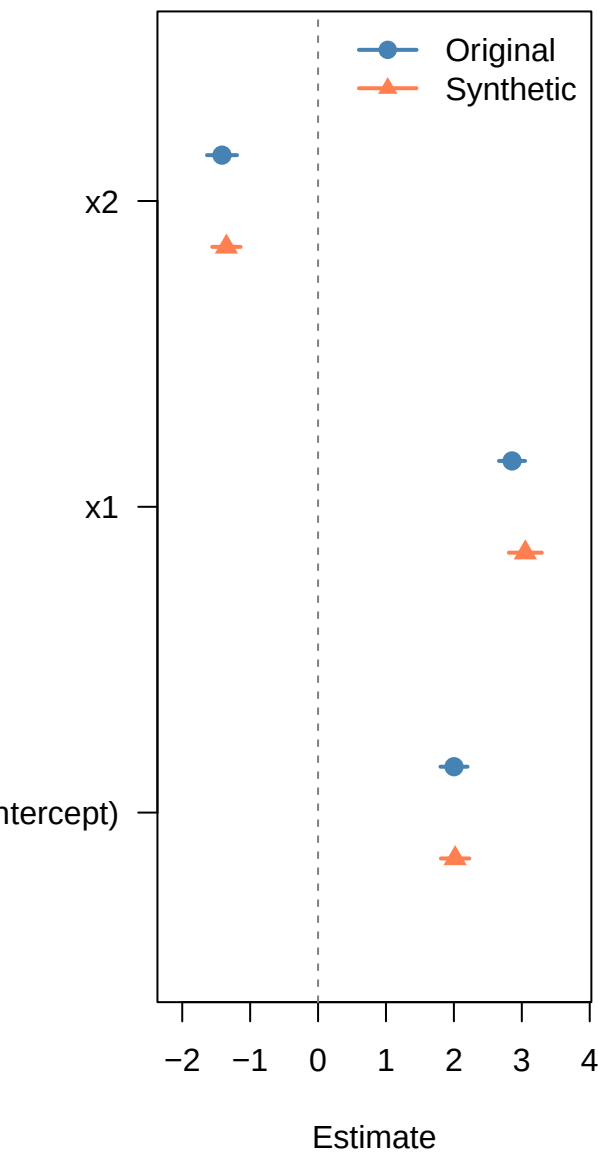


CI Overlap per Coefficient

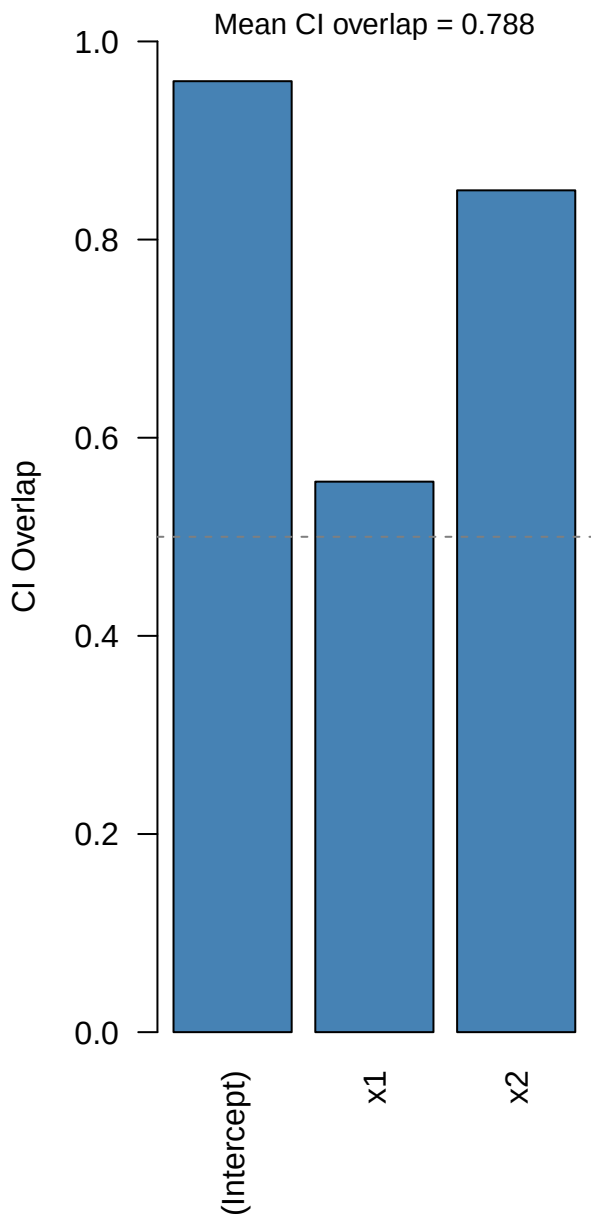
Mean CI overlap = 0.788



Regression Coefficients



CI Overlap per Coefficient



KL divergence

PC1

principal component

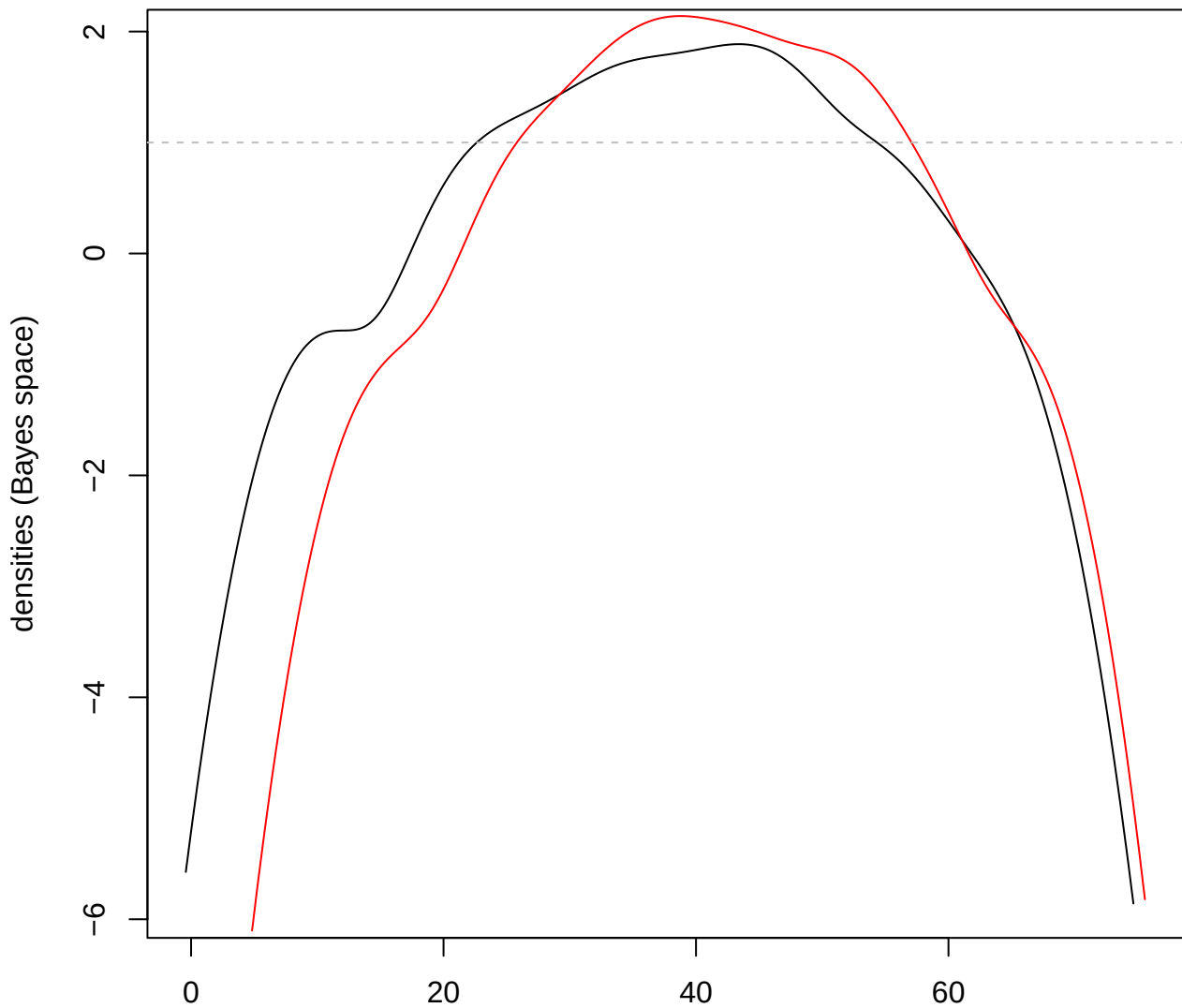
PC2

JS divergence

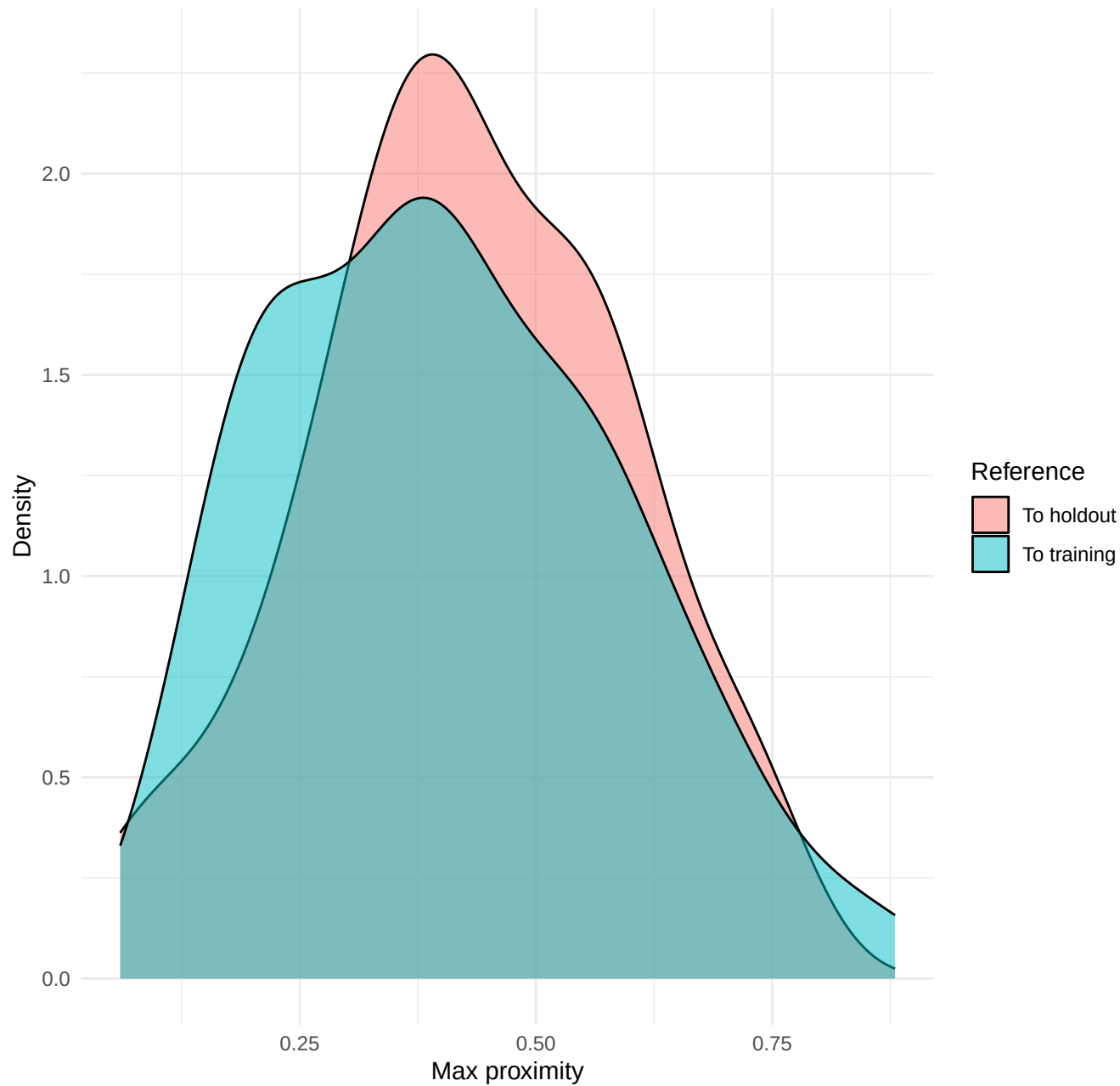
PC1

principal component

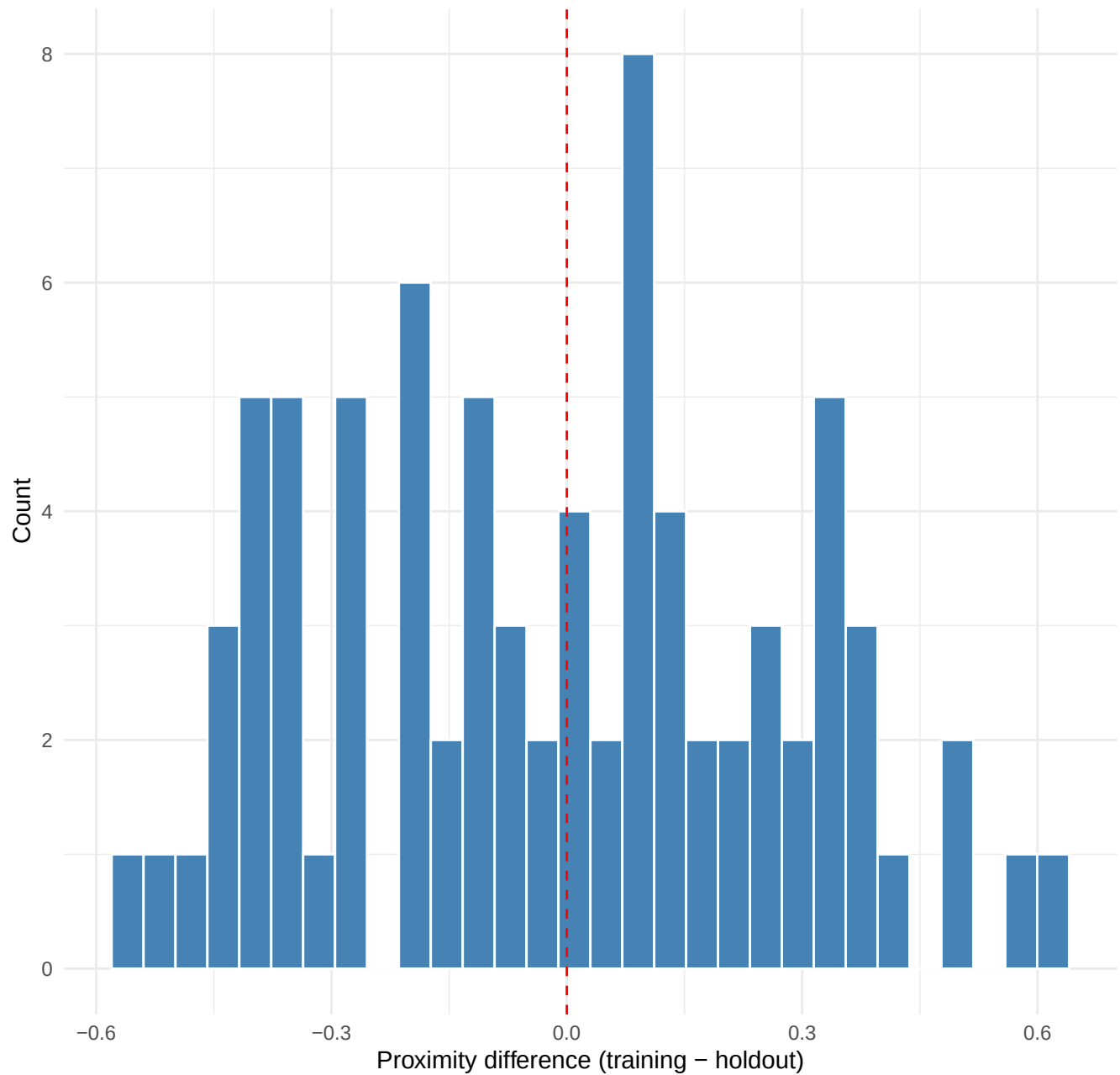
PC2



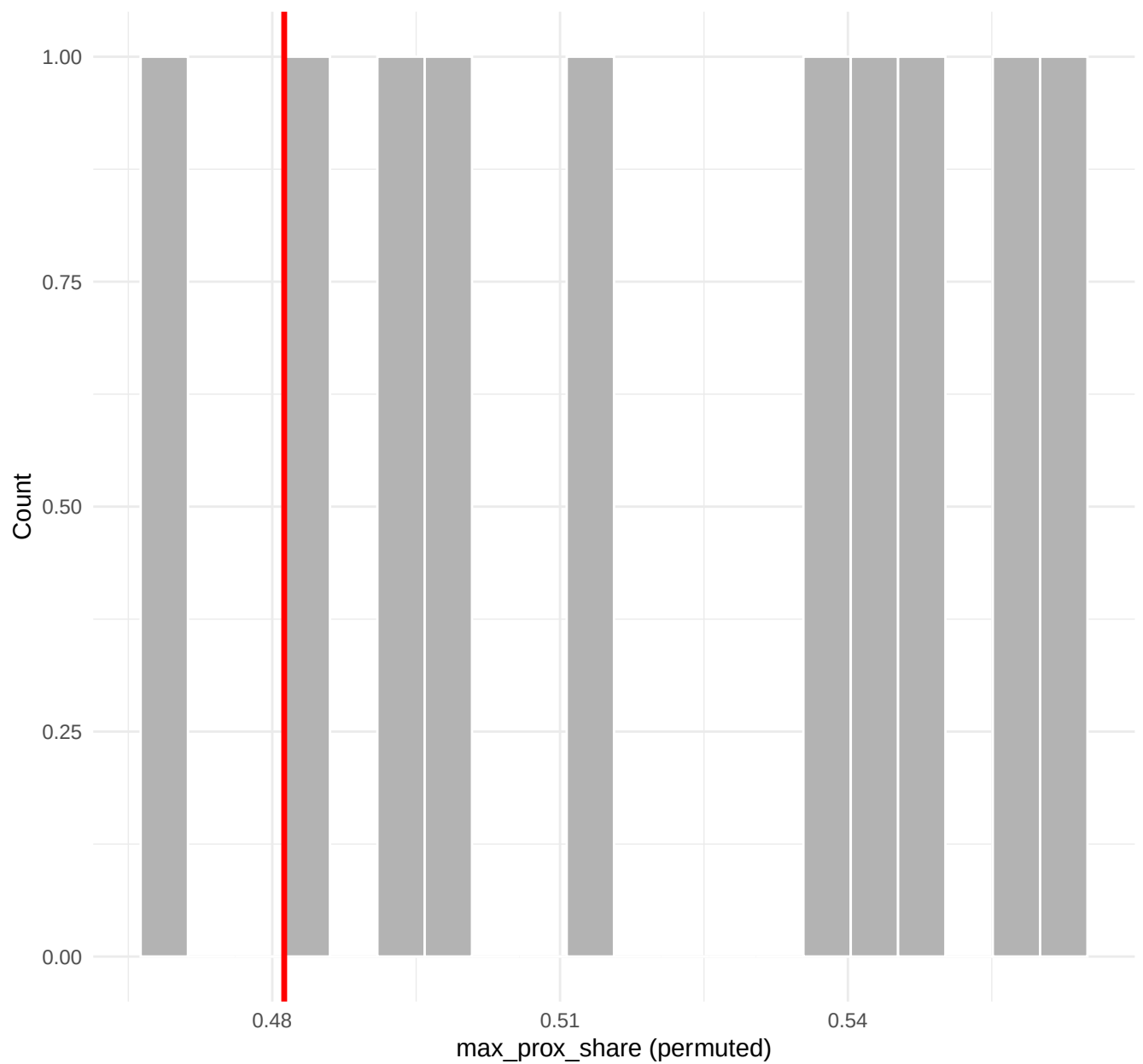
Max Proximity: Training vs Holdout



Per-Record Max Proximity Difference



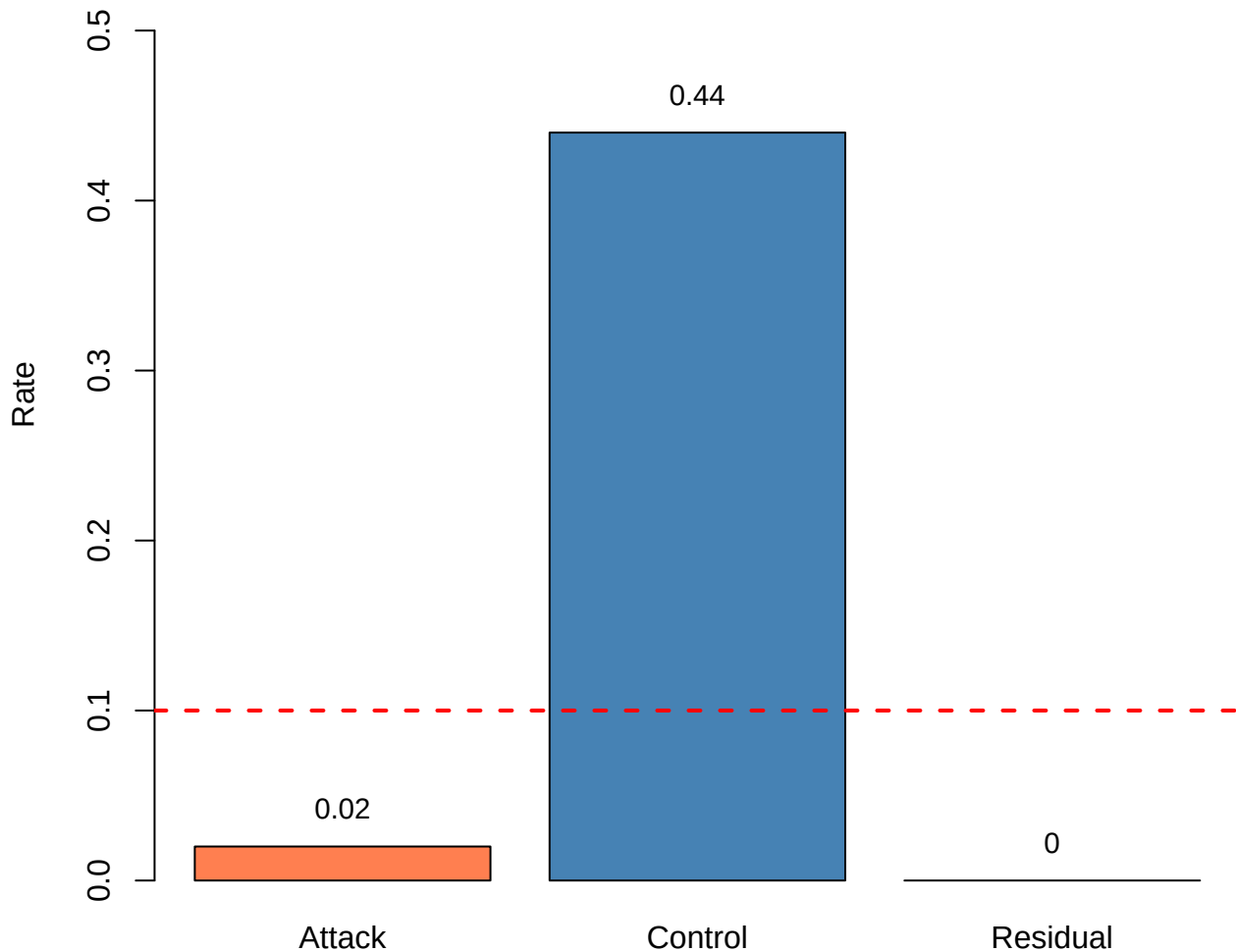
Null Distribution of Max Proximity Share



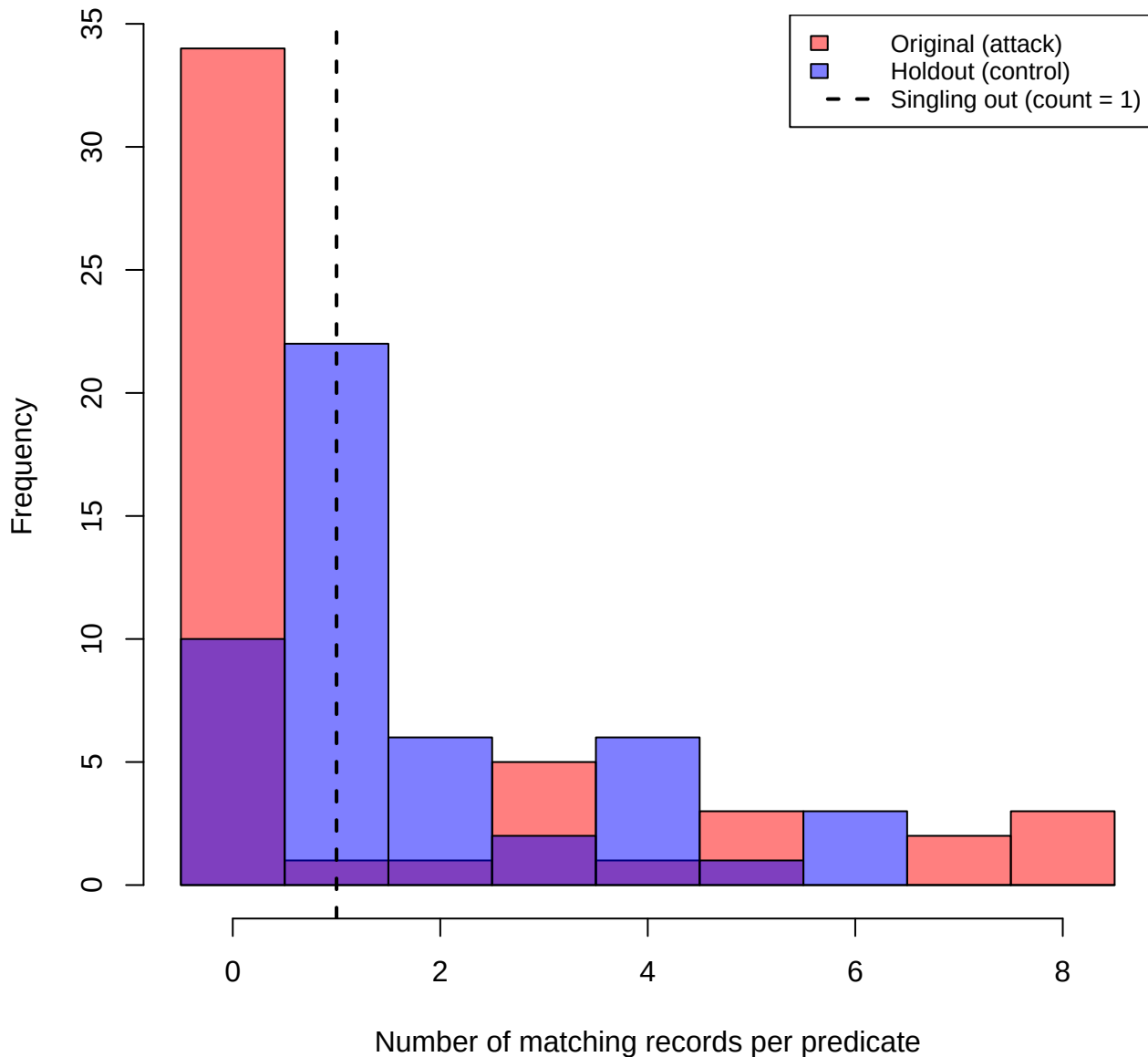
Observed = 0.481, $p = 0.909$

Singling Out Risk (50 predicates)

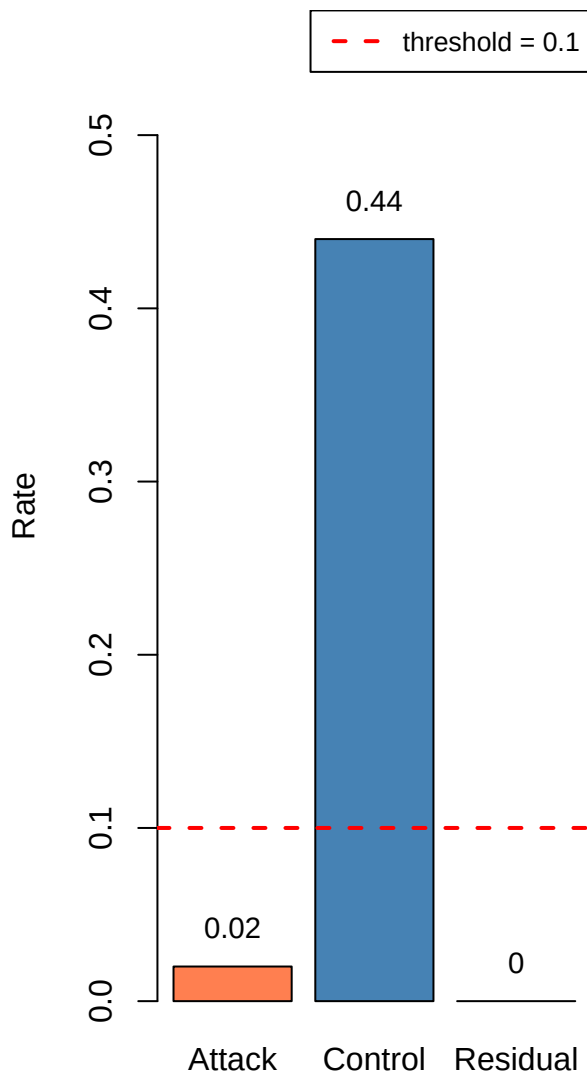
- - threshold = 0.1



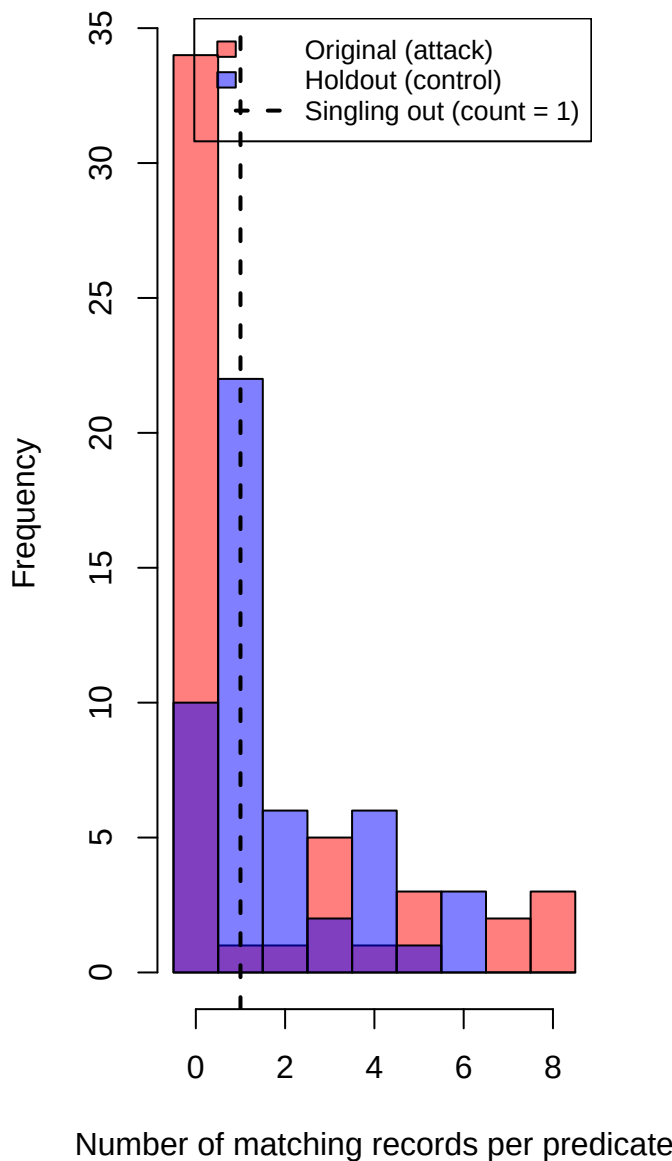
Match Count Distribution



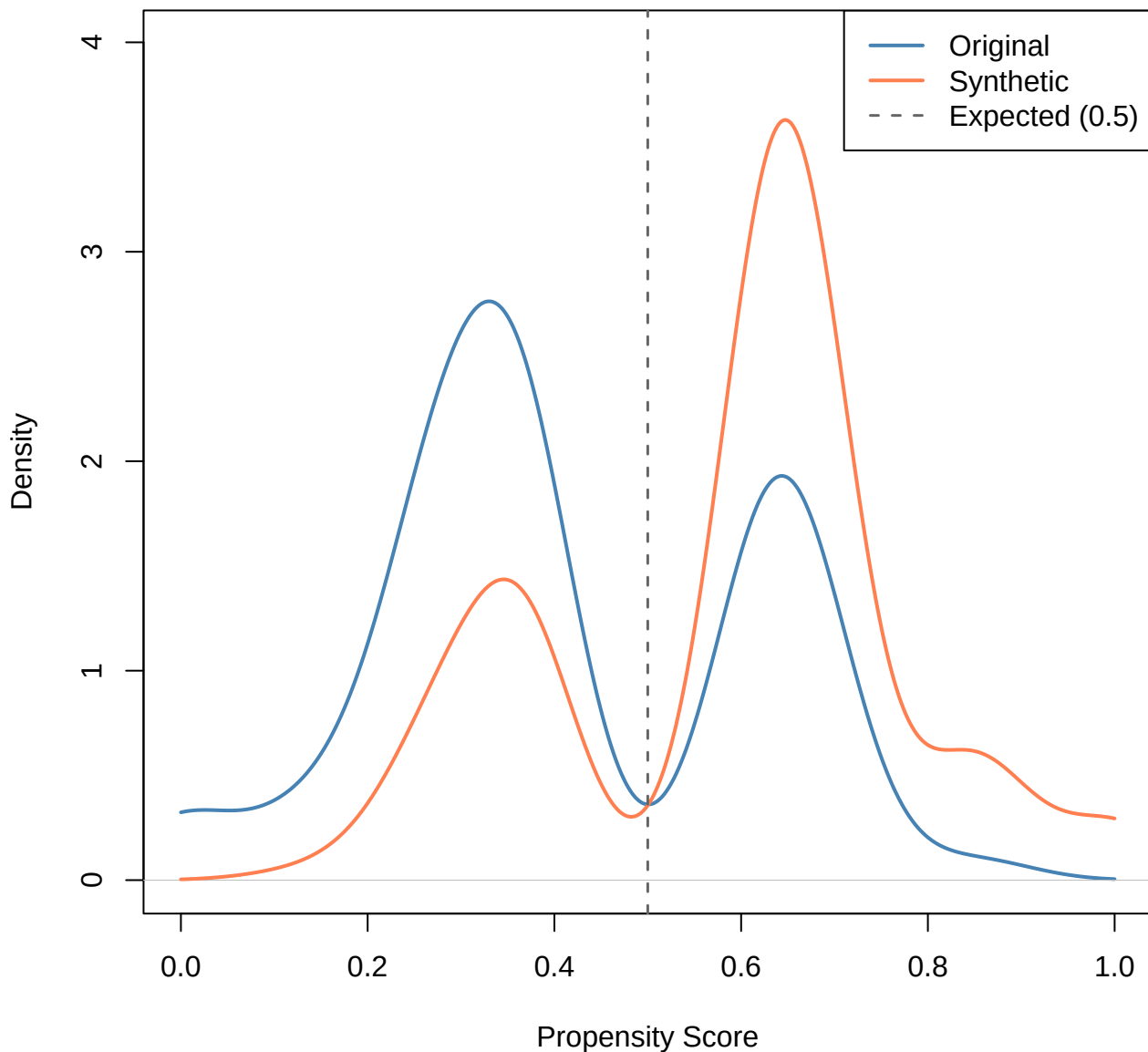
**Singling Out Risk
(50 predicates)**



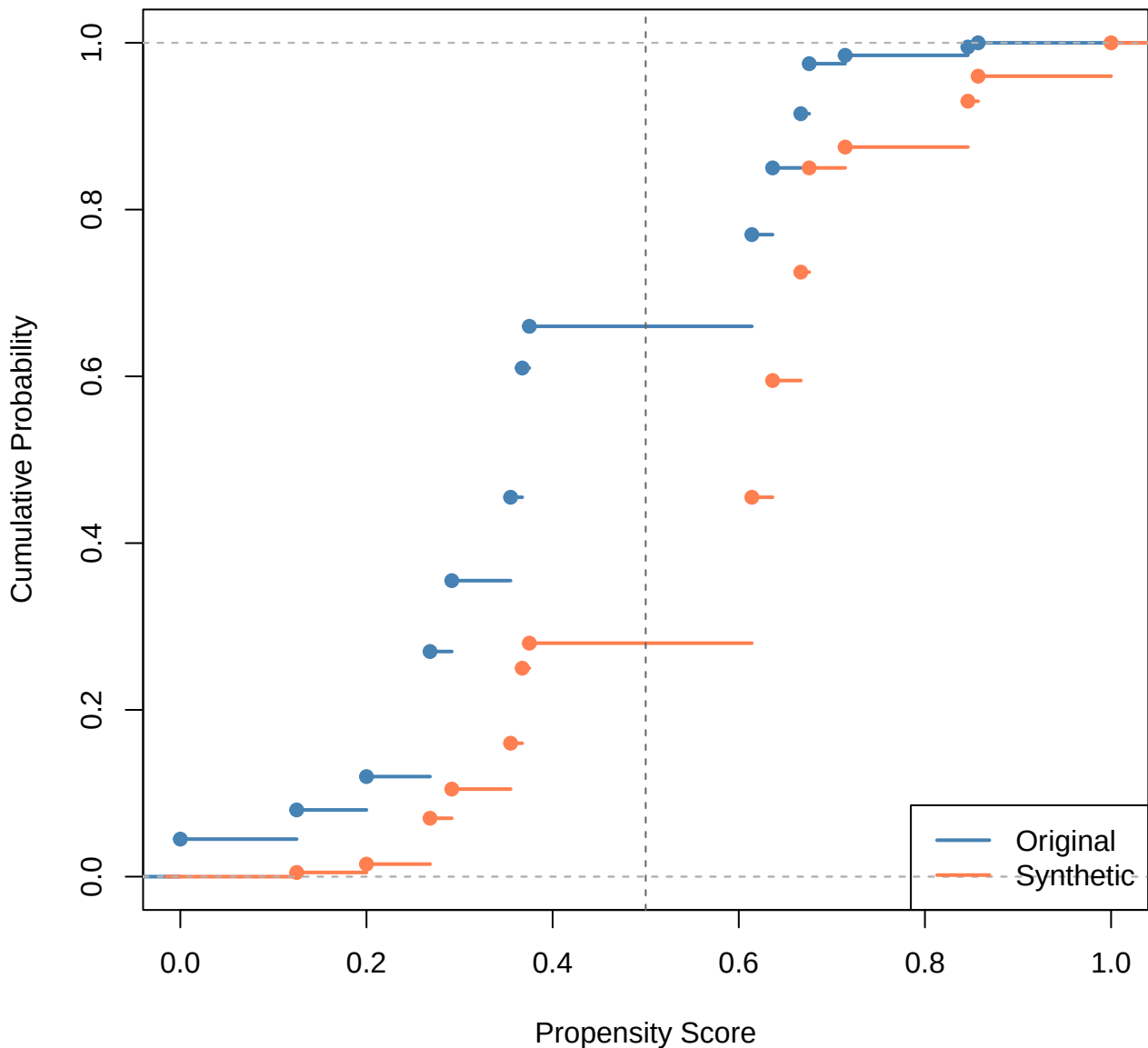
Match Count Distribution



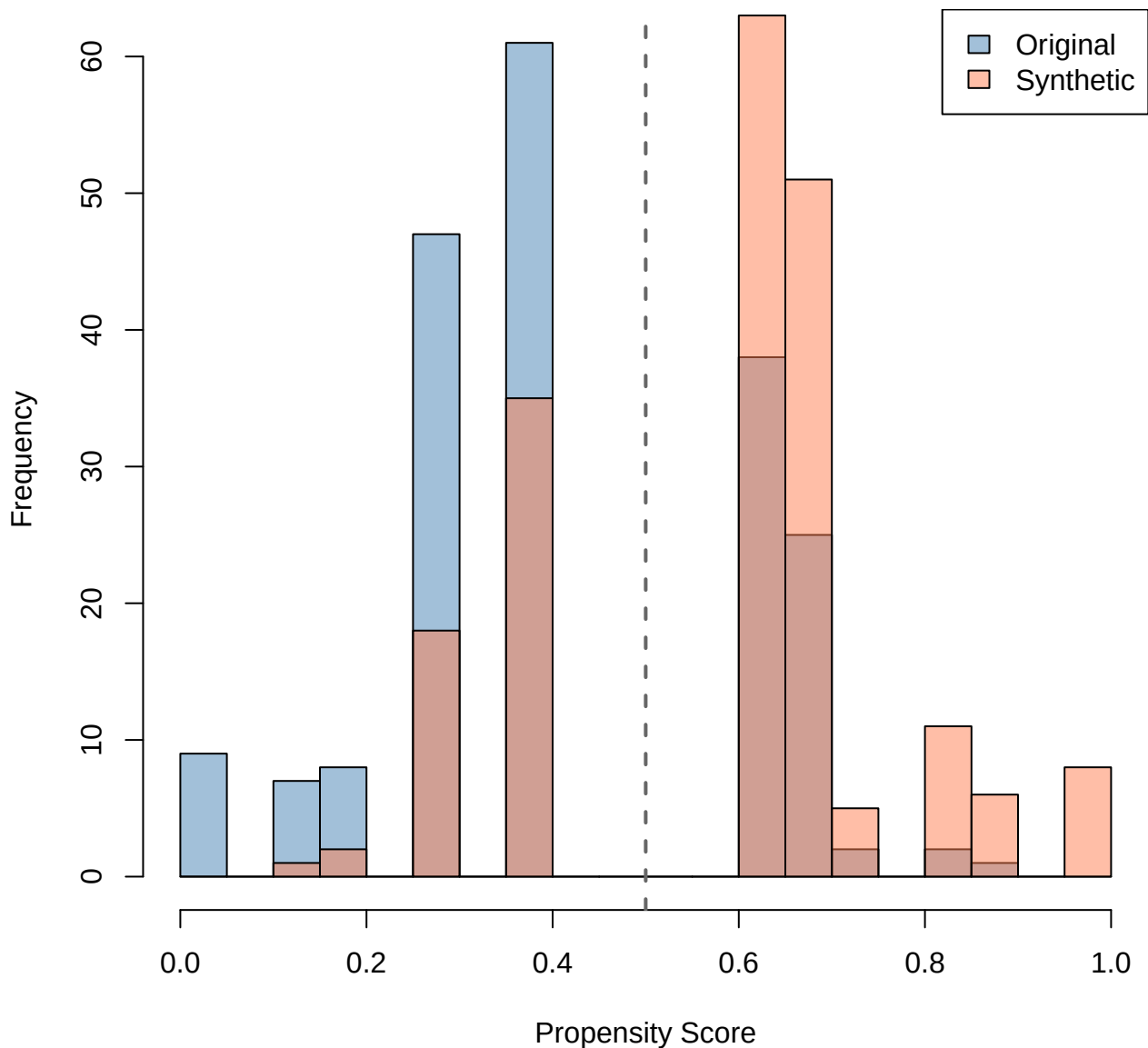
Propensity Score Densities



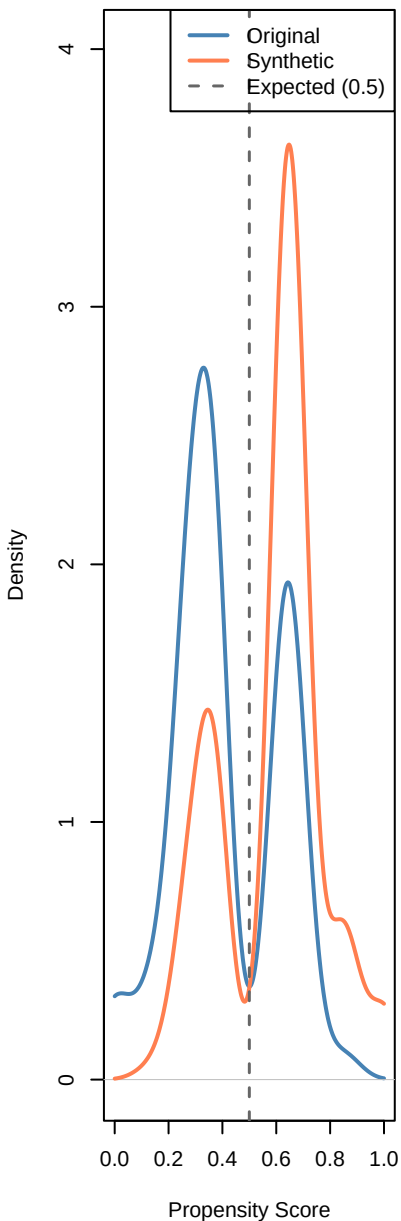
ECDF Comparison (SPECKS = 0.38)



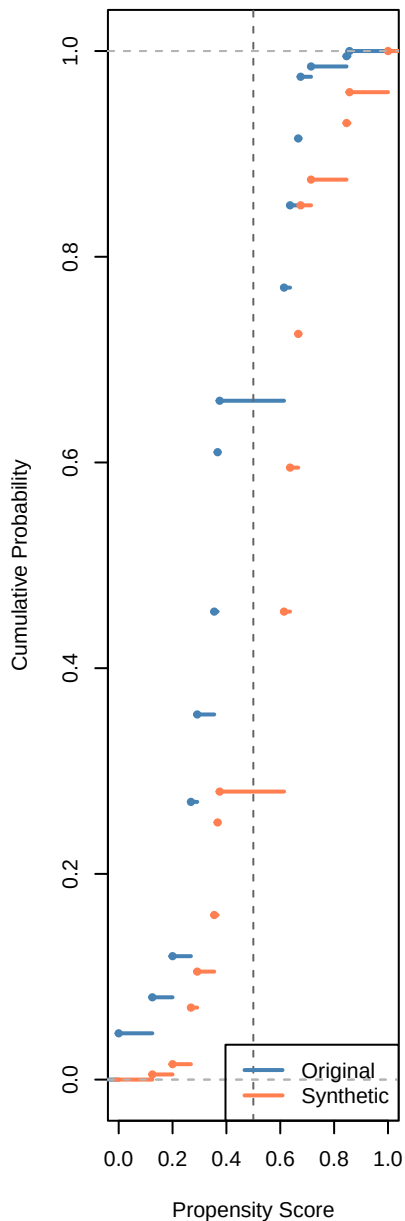
Propensity Score Distributions



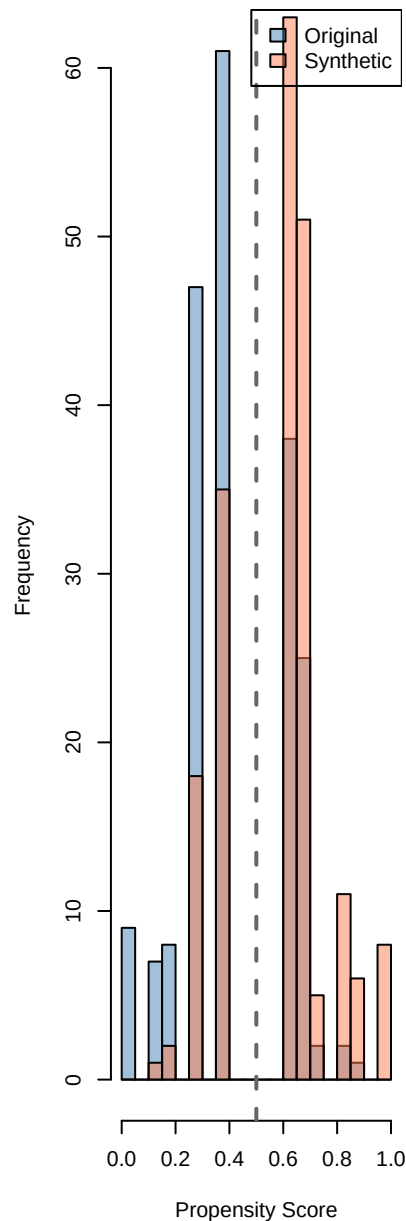
Propensity Score Densities



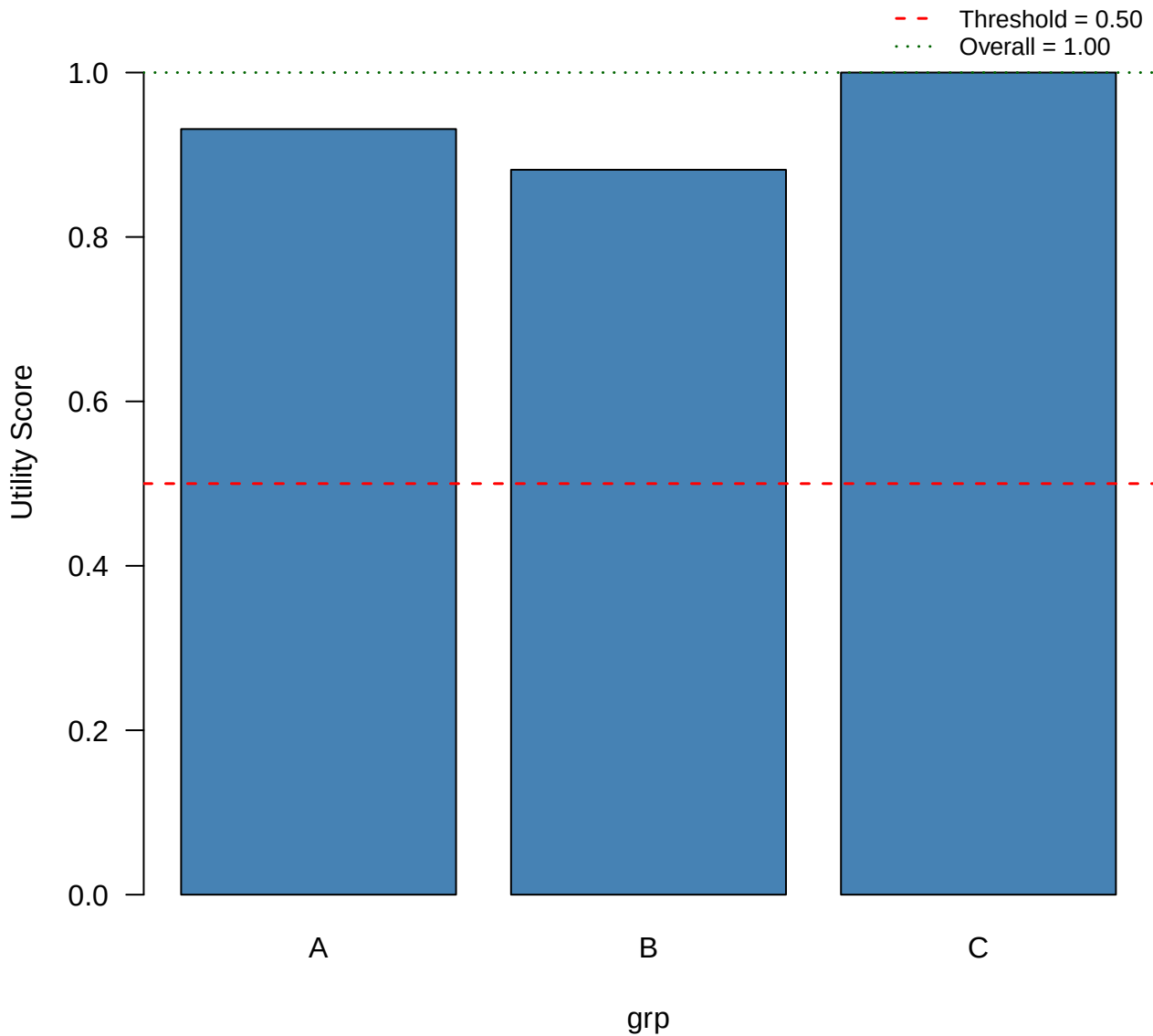
ECDF Comparison (SPECKS = 0.3)



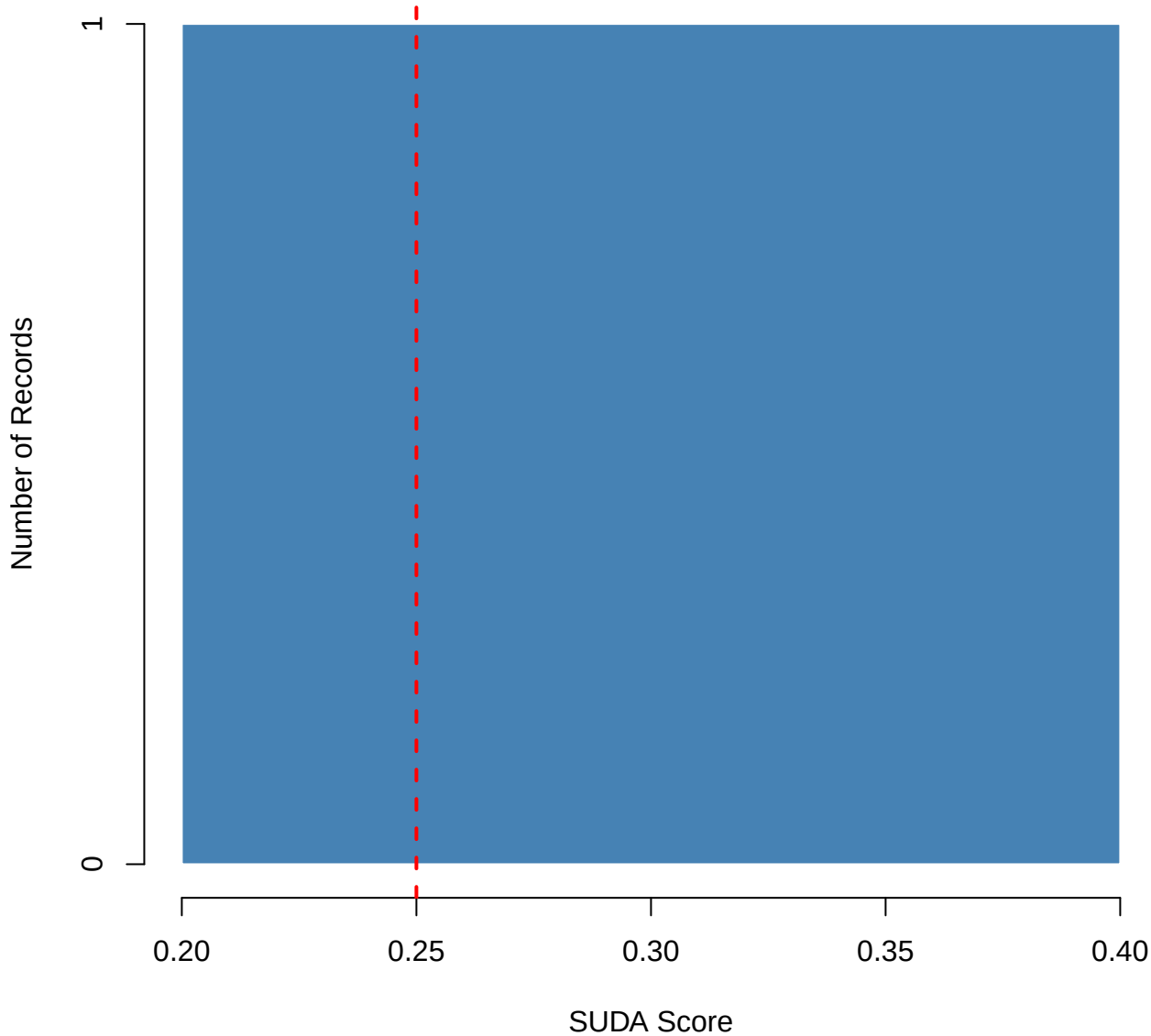
Propensity Score Distributions



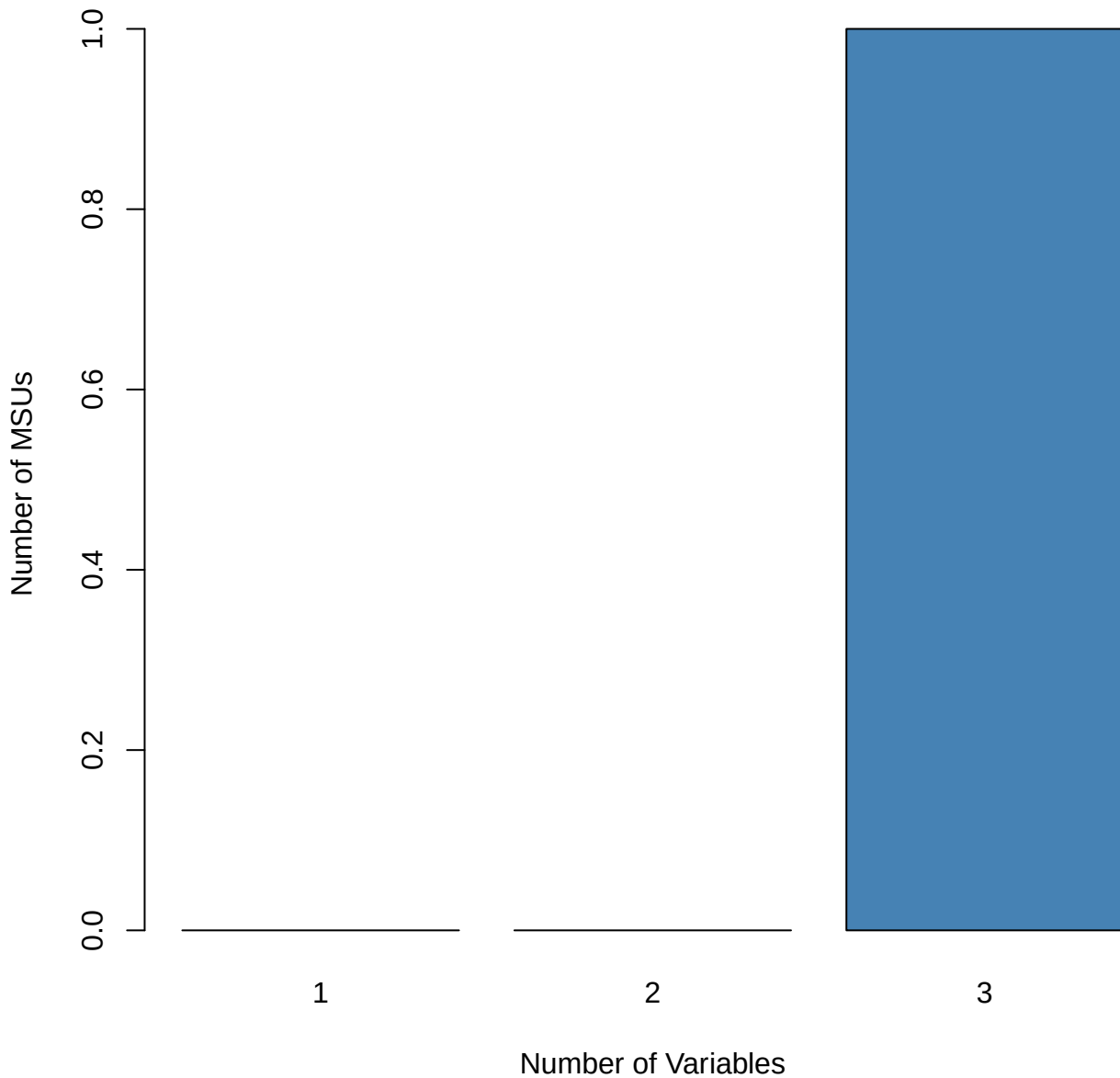
Subgroup Utility by grp



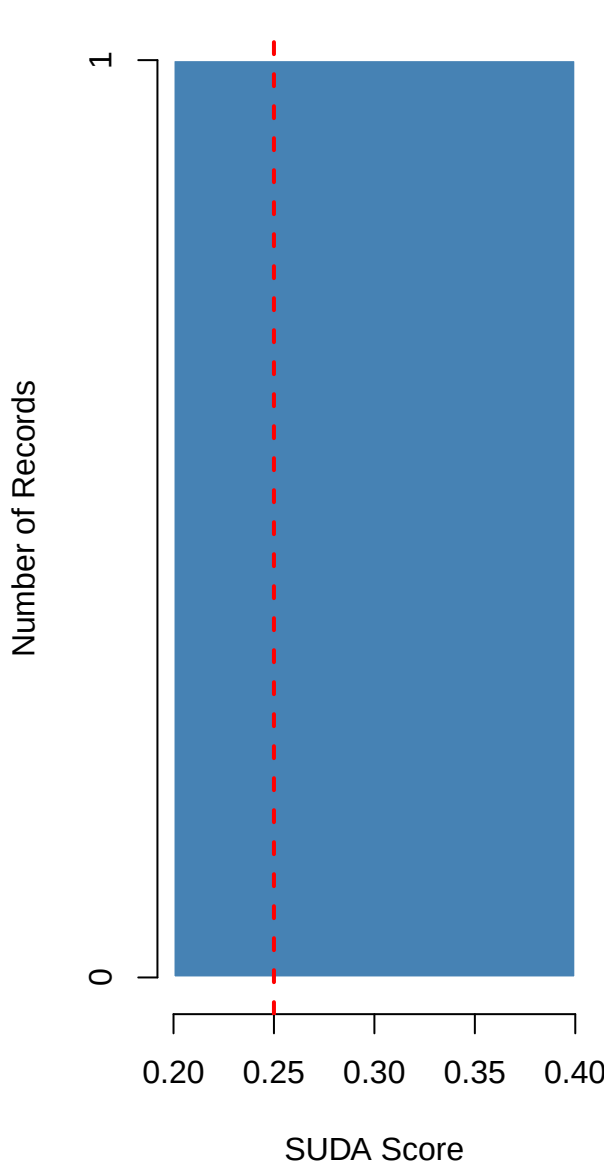
Distribution of SUDA Scores (non-zero)



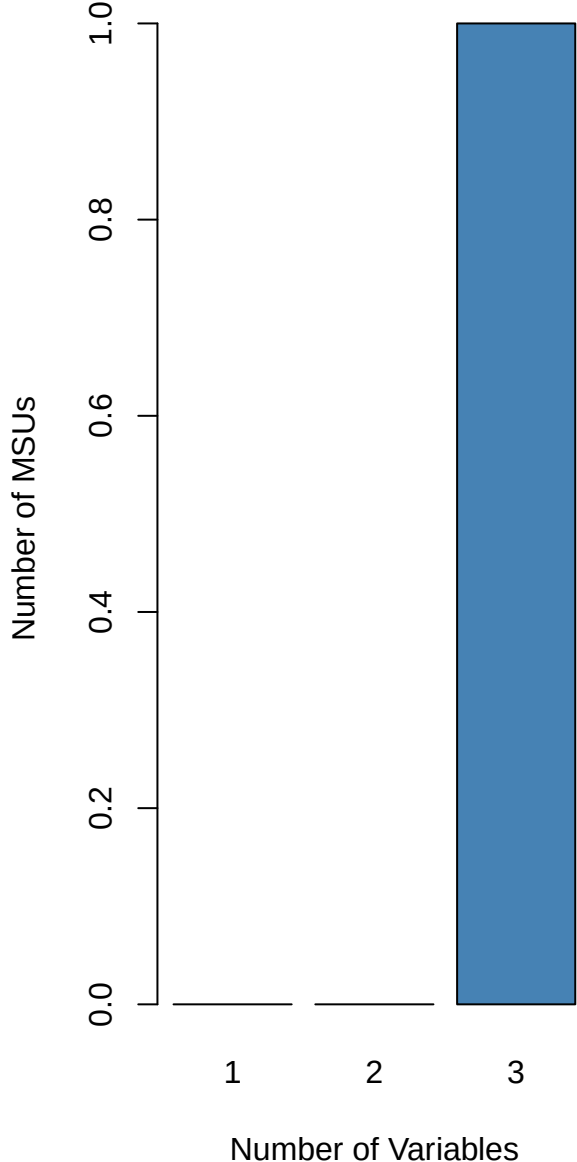
MSUs by Variable Subset Size



Distribution of SUDA Scores (non-zero)



MSUs by Variable Subset Size



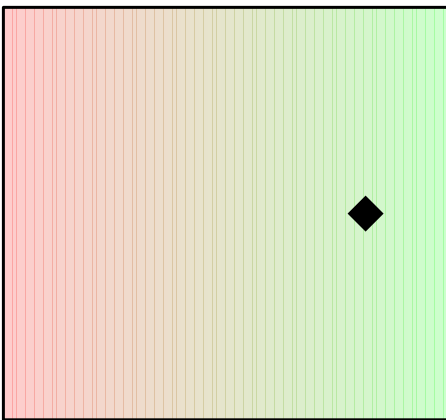
No MSUs Found

No records with SUDA > 0

No MSUs Found

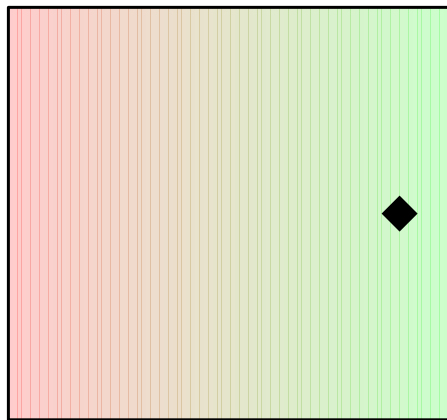
No MSUs detected

a



QQ div = 0.205

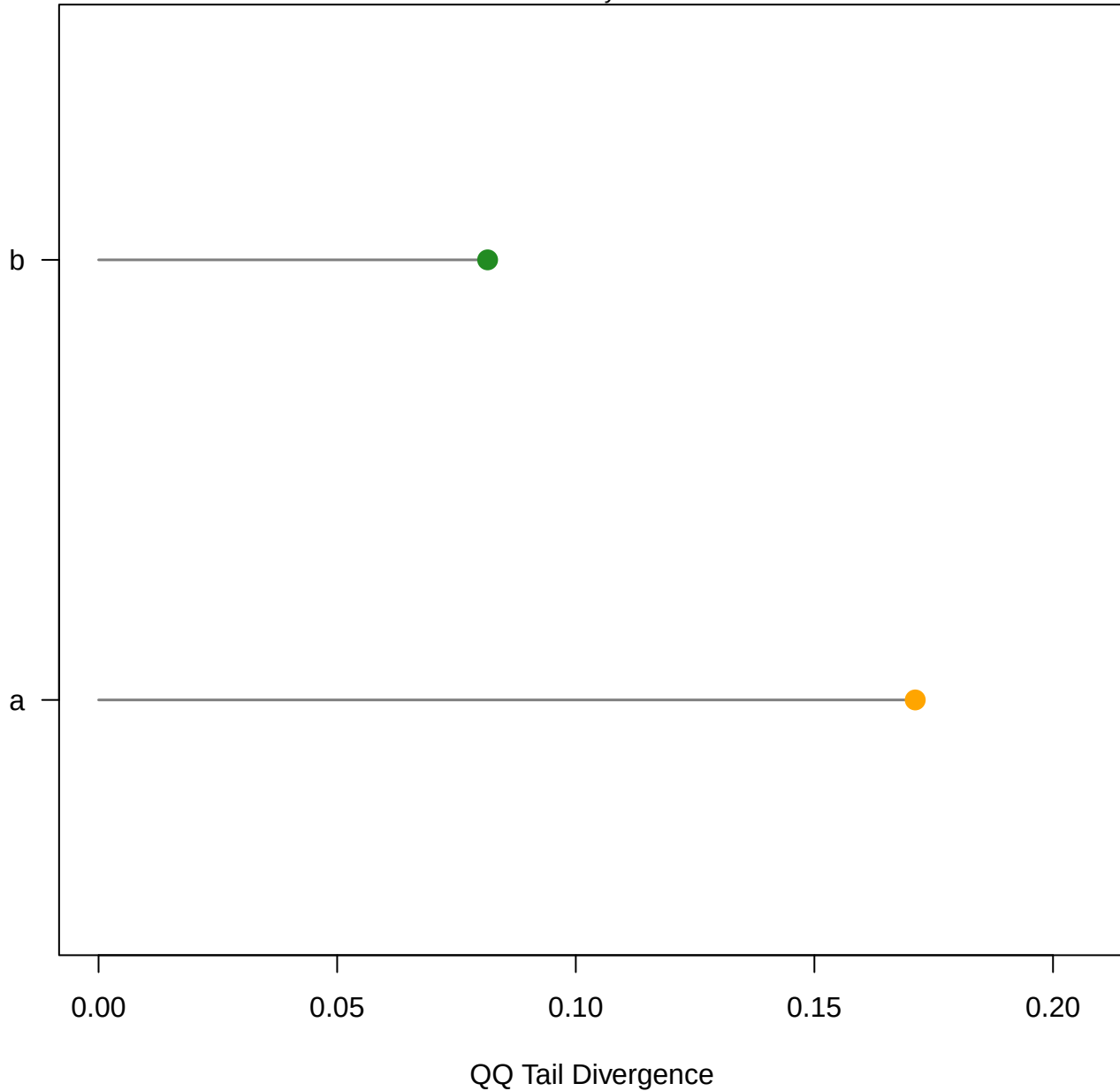
b



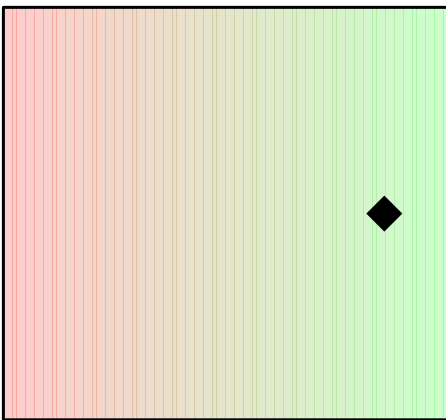
QQ div = 0.128

Tail Fidelity (percentile = 95, tails = both)

Overall utility = 0.881

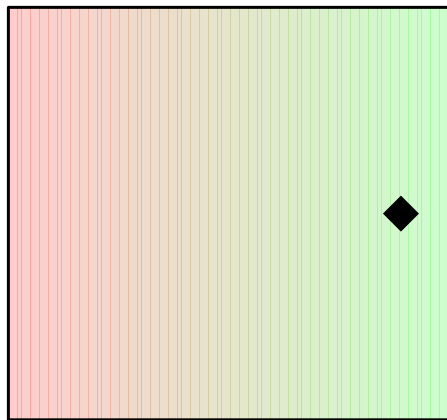


a



QQ div = 0.154

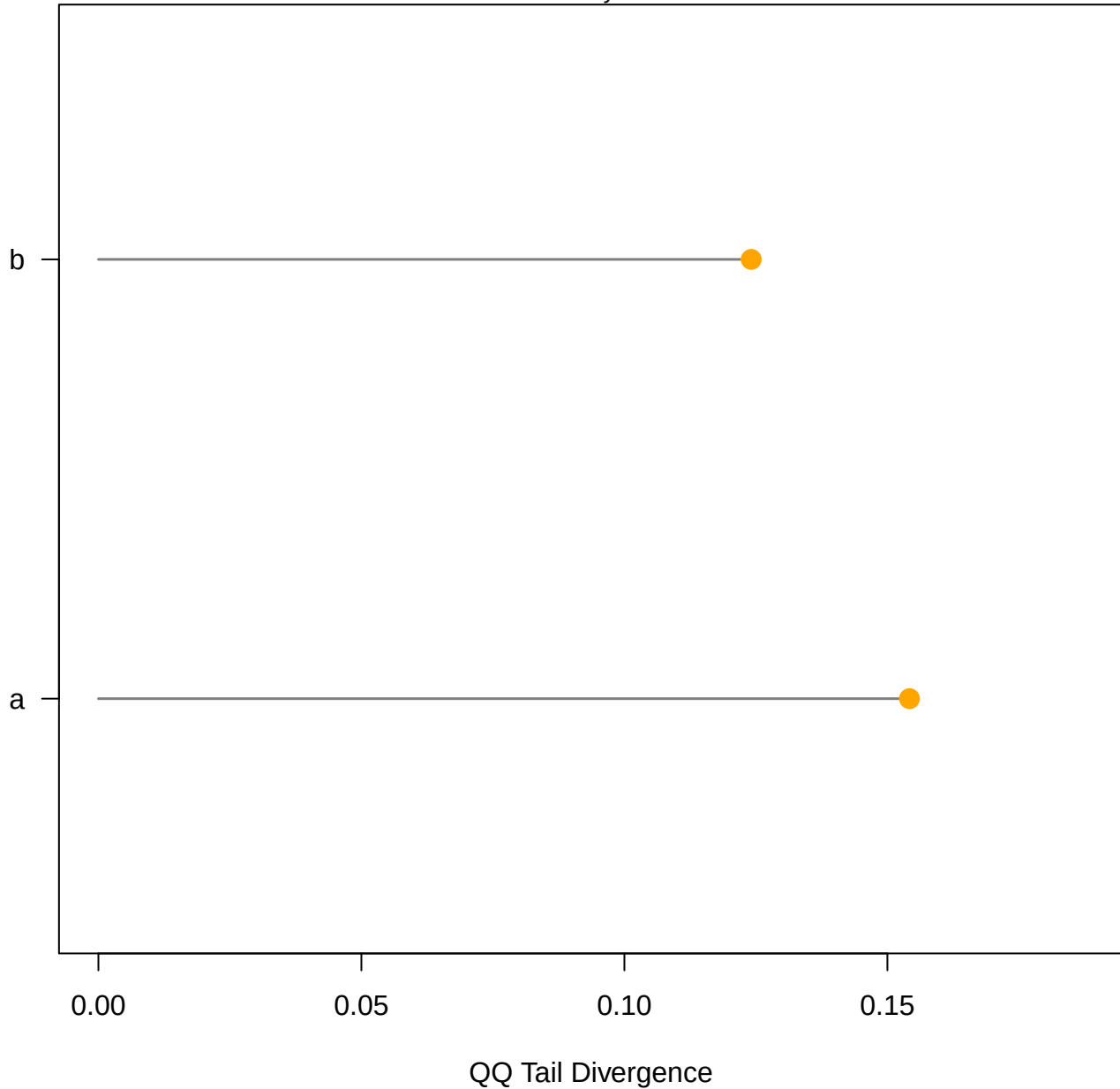
b



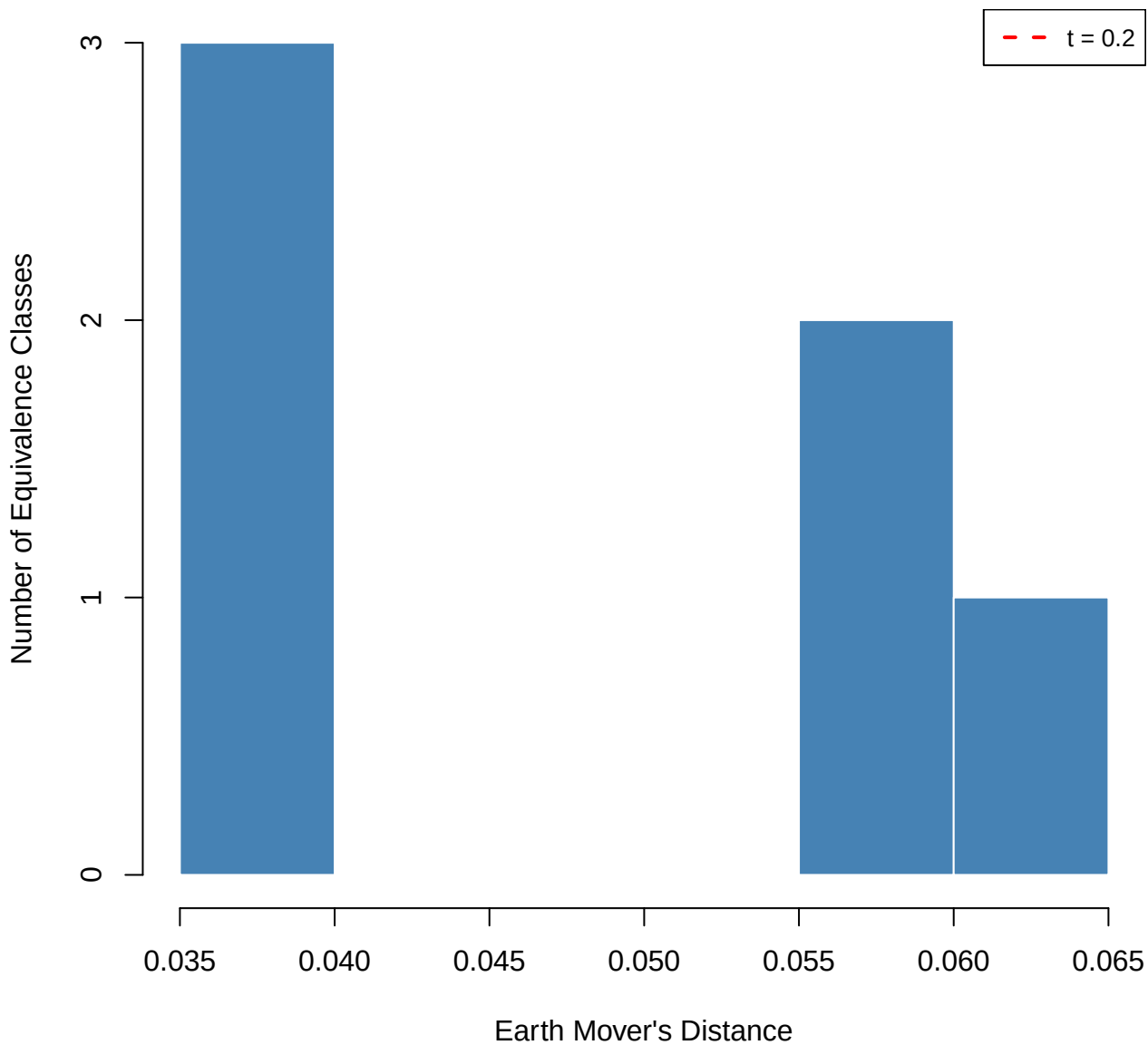
QQ div = 0.124

Tail Fidelity (percentile = 95, tails = both)

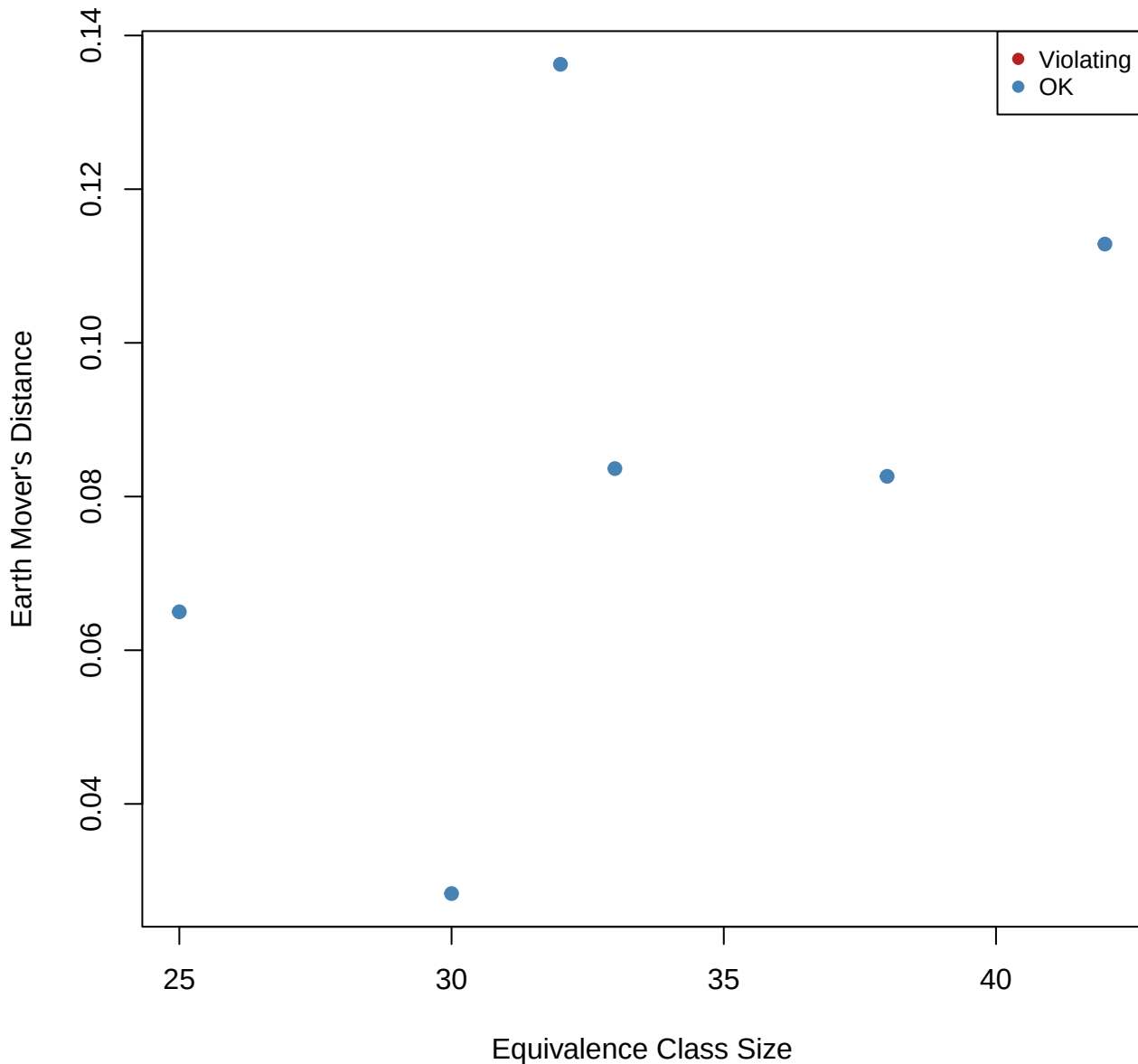
Overall utility = 0.870



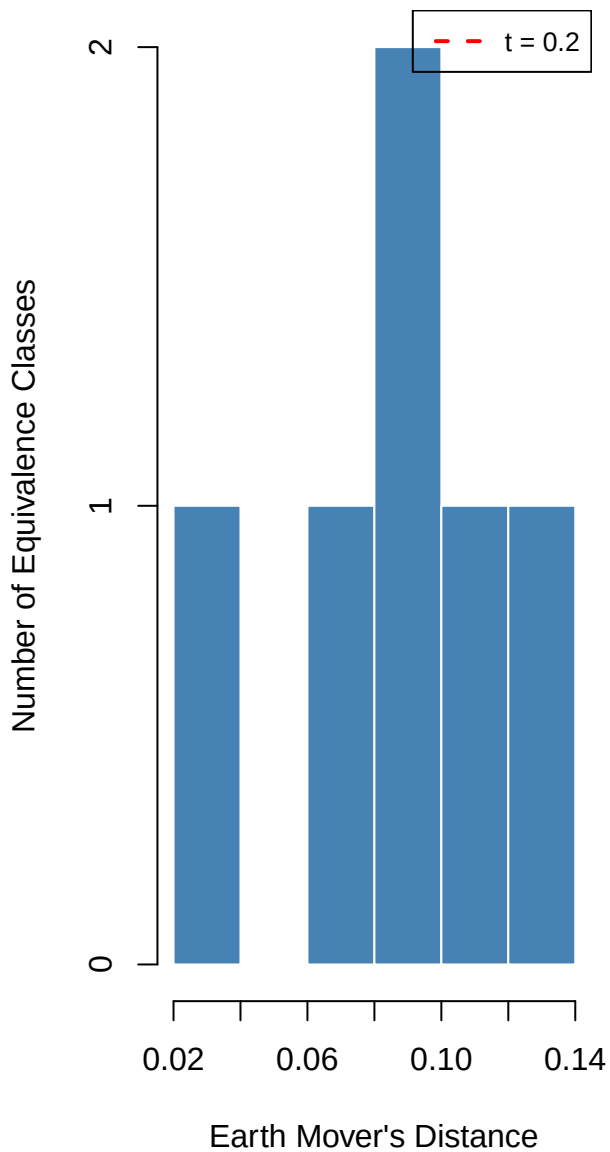
EMD Distribution (t = 0.2)



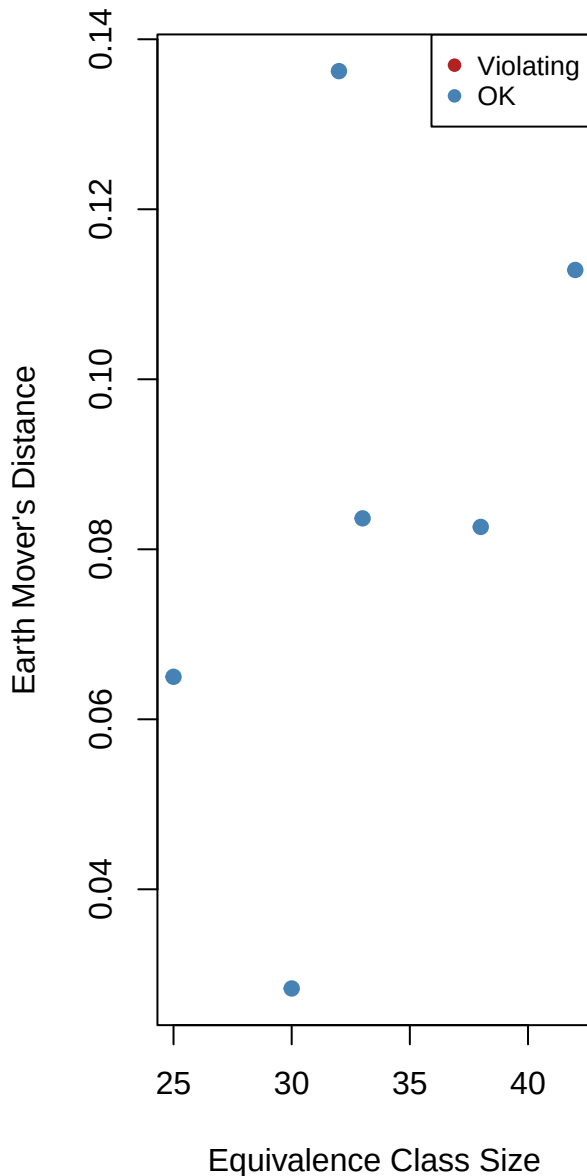
EC Size vs EMD



EMD Distribution (t = 0.2)

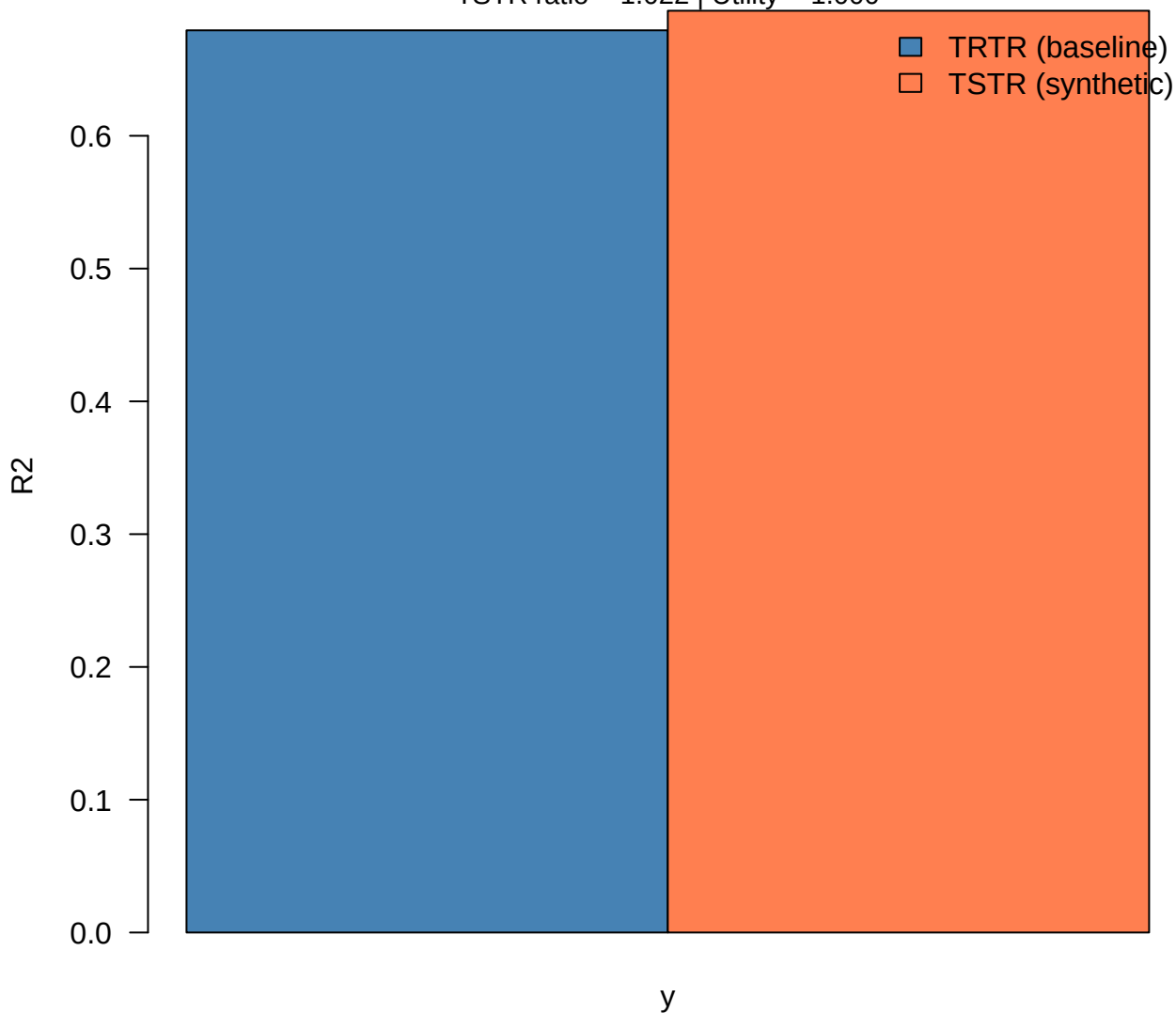


EC Size vs EMD



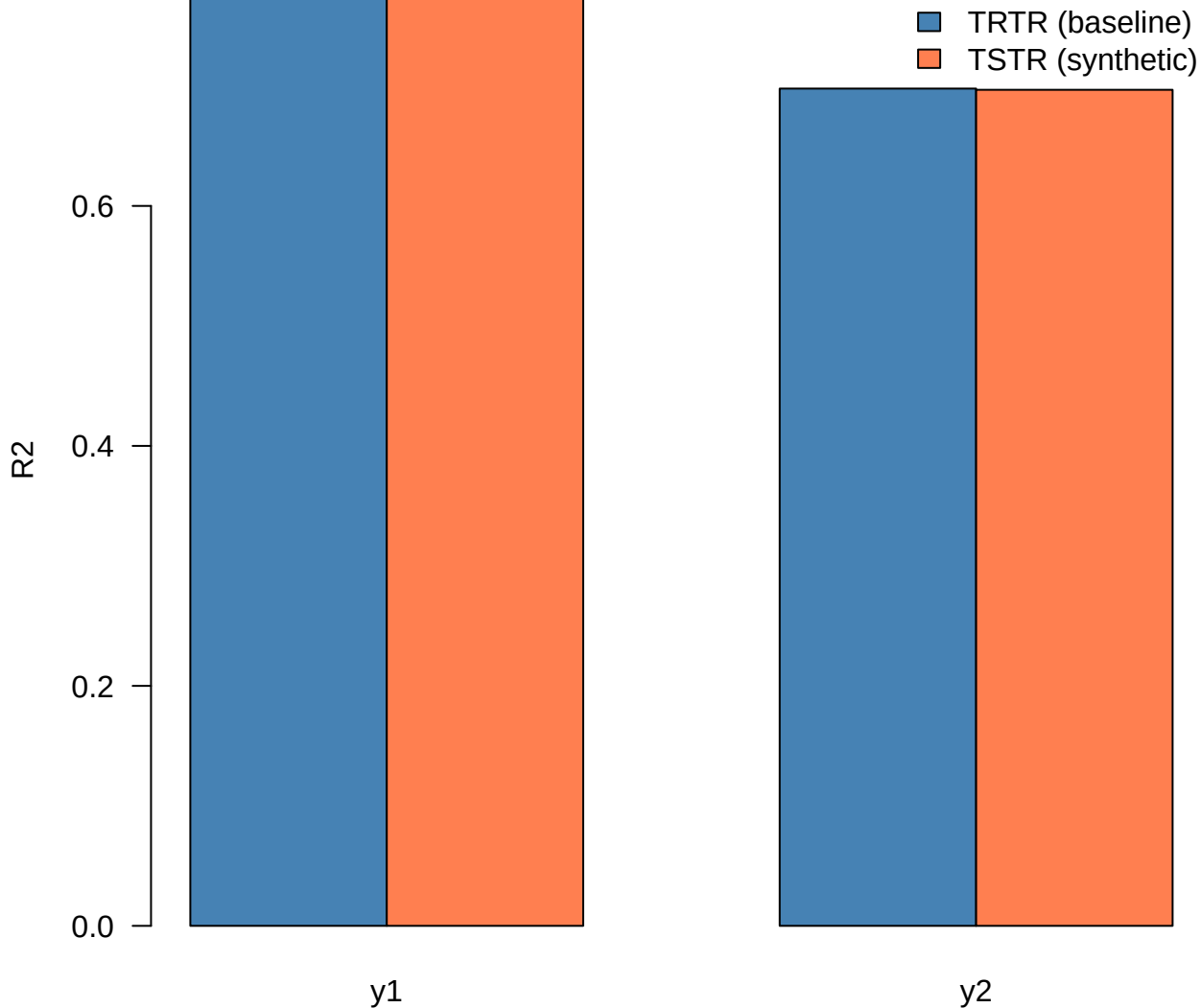
TRTR vs TSTR Performance

TSTR ratio = 1.022 | Utility = 1.000

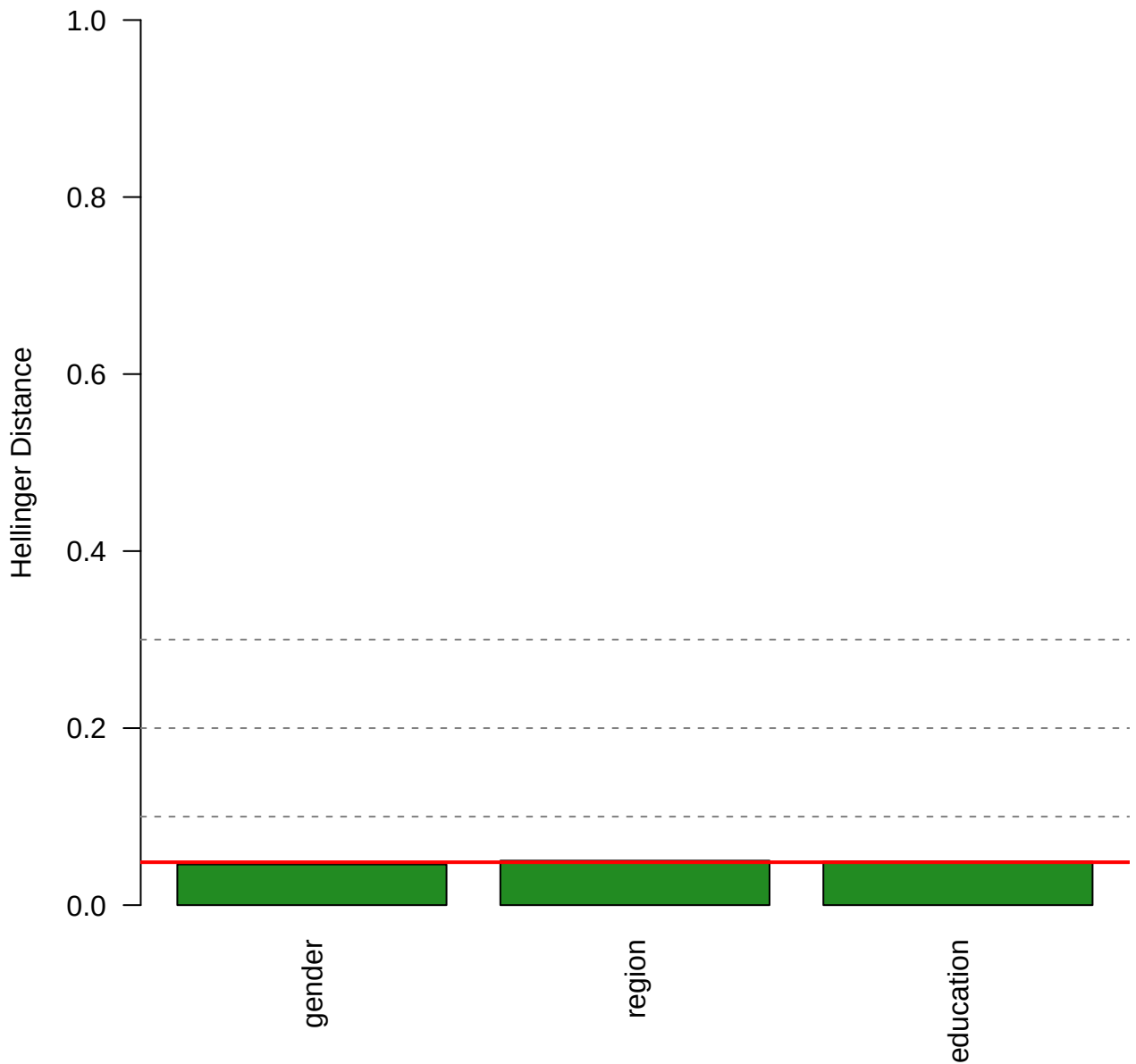


TRTR vs TSTR Performance

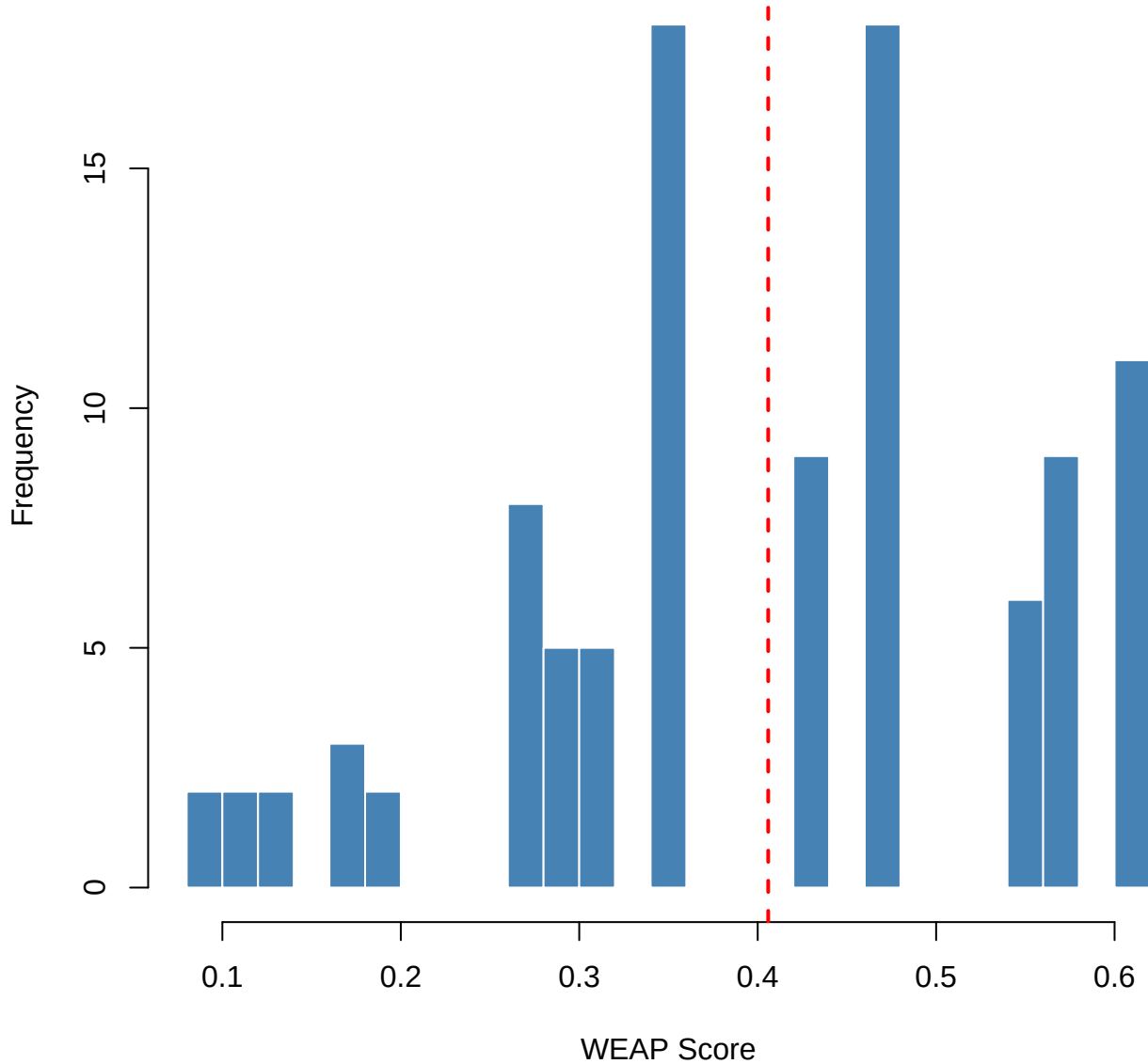
TSTR ratio = 1.004 | Utility = 1.000



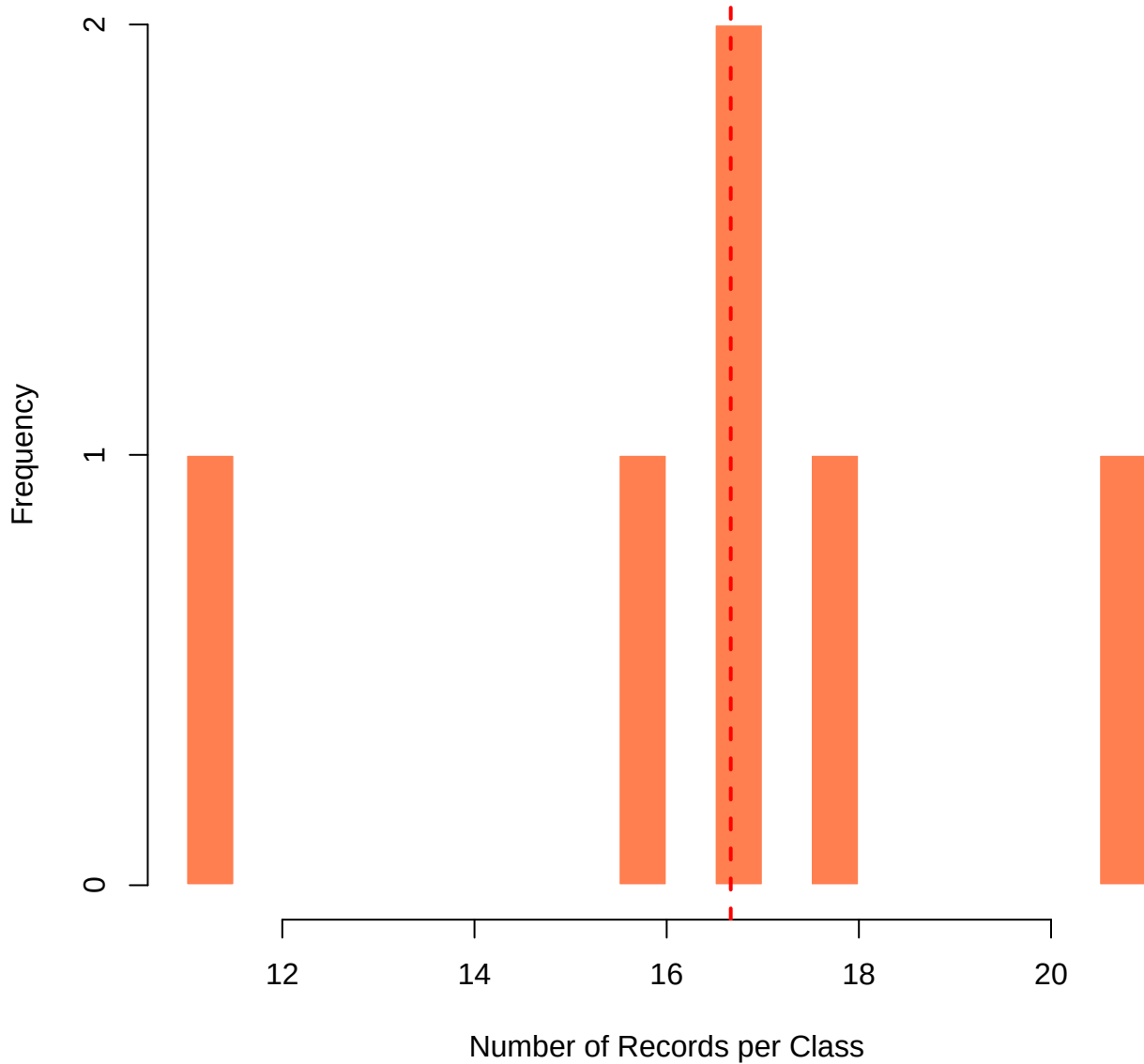
Hellinger Distance by Variable



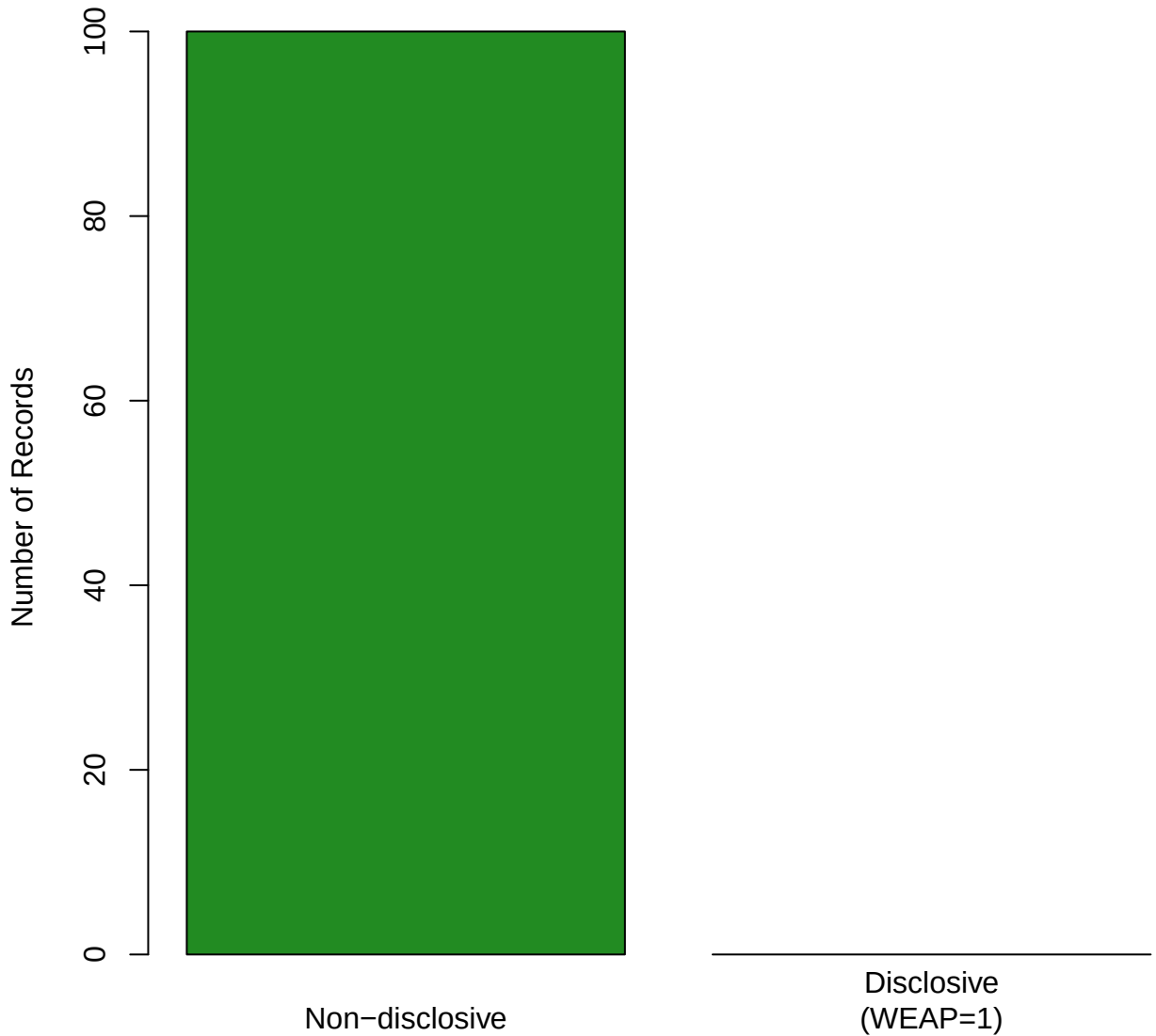
Distribution of WEAP Scores



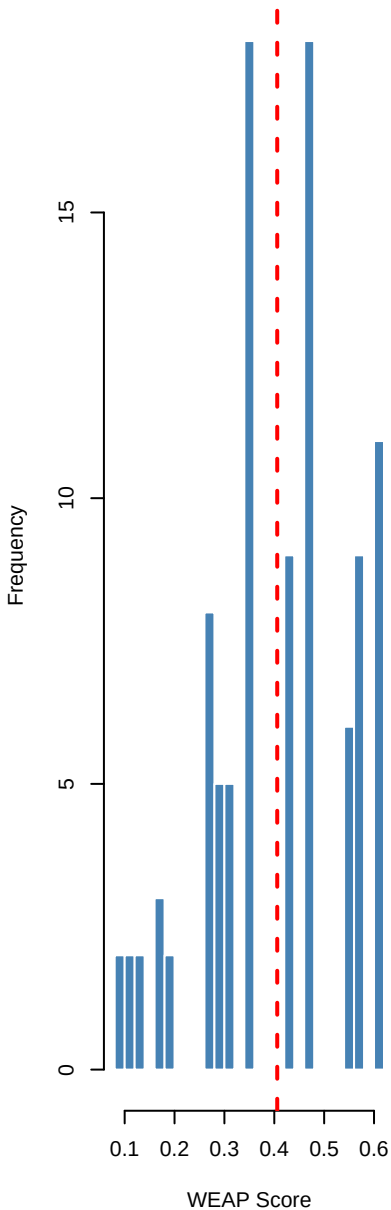
Equivalence Class Sizes



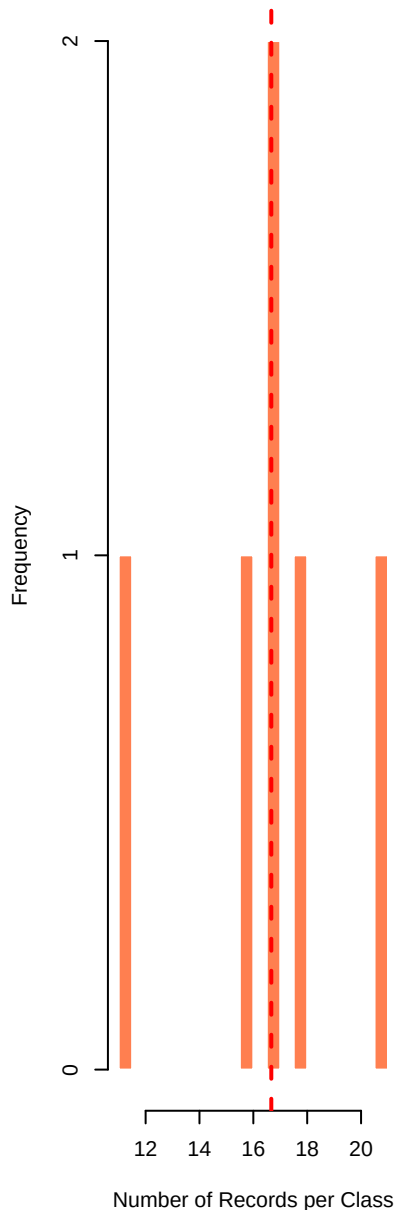
Record Disclosure Status



Distribution of WEAP Scores



Equivalence Class Sizes



Record Disclosure Status

