

Assignment-Control Plots: A Visualization tool for Causal Inference Studies

Rachael C. Aikens, Michael Baiocchi

1 Set-up

The results that follow depict several simulated datasets. The primary generative model for these is based on Aikens, Greaves, and Baiocchi (2020) and is specified as follows:

$$\begin{aligned}X_i &\sim \text{Normal}(0, I_{10}) \\T_i &\sim \text{Bernoulli}\left(\frac{1}{1 + \exp(\phi(X_i))}\right) \\Y_i(0) &= \Psi(X_i) + \epsilon_i \\ \epsilon_i &\sim N(0, 1)\end{aligned}$$

where $\phi(X)$ and $\Psi(X)$ represent the true propensity and prognostic score functions, given by

$$\begin{aligned}\phi(X_i) &= c_1 X_{i1} - c_0 \\ \Psi(X_i) &= \rho X_{i1} + \sqrt{(1 - \rho^2)} X_{i2}.\end{aligned}$$

Where c_1 , c_0 , and ρ are constants. In particular, the form for the prognostic function above guarantees that $\rho = \text{Corr}(\phi(X), \Psi(X))$.

2 Results

2.1 Figure 1: Assignment-Control Plots and Data Diagnostics

A first use for the assignment-control plot is as an informal data diagnostic. Researchers running observational studies often want a basic understanding of the baseline variation between the treatment and control groups before they begin. Plotting the standardized mean differences between the treatment groups (love plots) is a common starting place to understand imbalances between groups, but when there are many covariates – some of which are irrelevant – it can be difficult to tell by eye which imbalances should be of most concern. Histograms of propensity score overlap can be an important diagnostic for checking that the treatment and control groups overlap in terms of their probability of treatment (since this condition is essential for many causal inference methodologies). However, the propensity score does not necessarily reflect all aspects of covariate balance which may be important.

```
## pdf
## 2
```

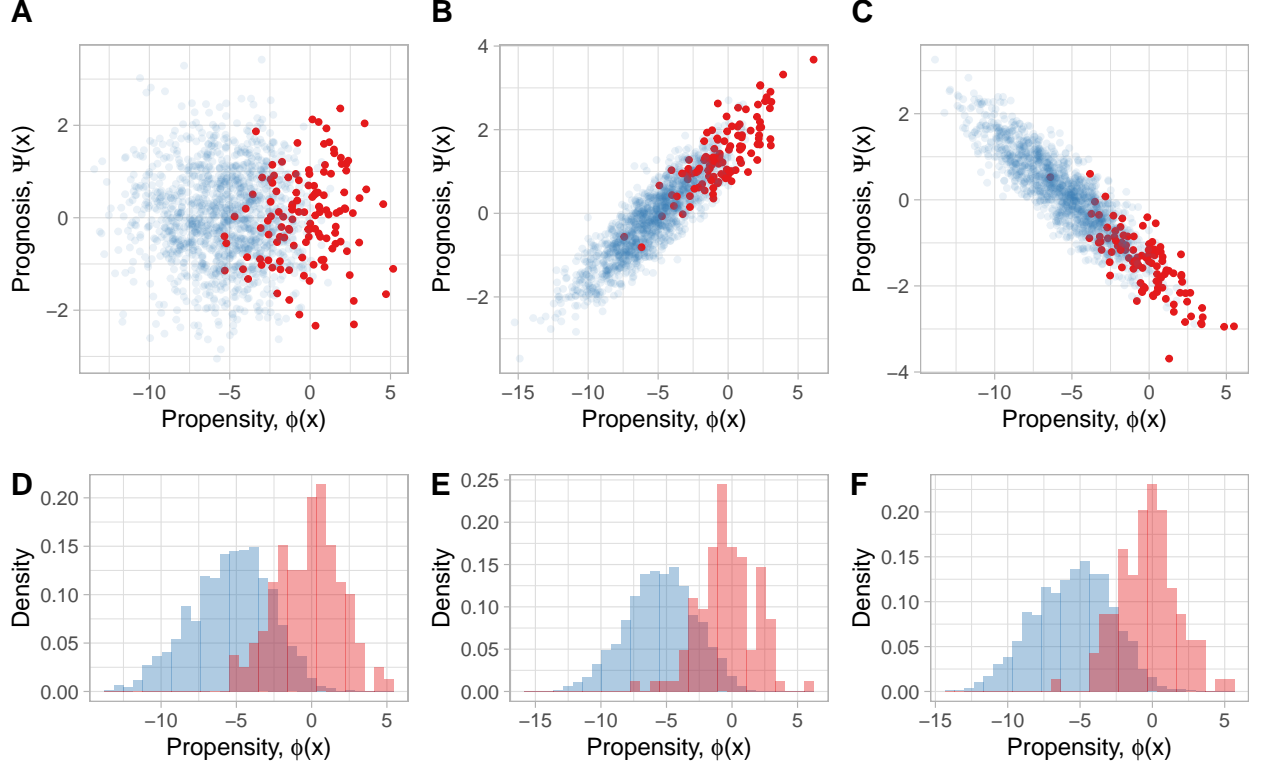
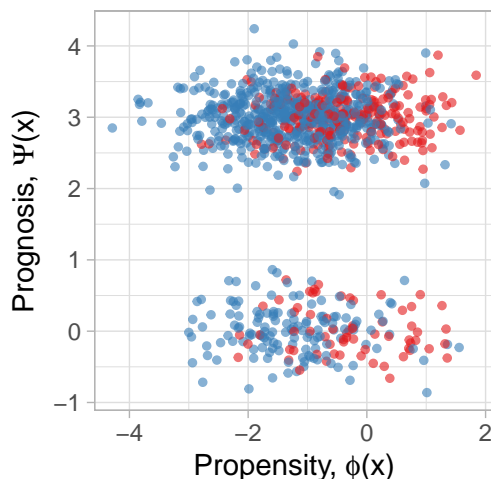


Figure 1: Assignment-Control plots (A-C) and propensity score density histograms (D-F) for three simulated observational datasets. Red points represent treated observations, blue points represent control.

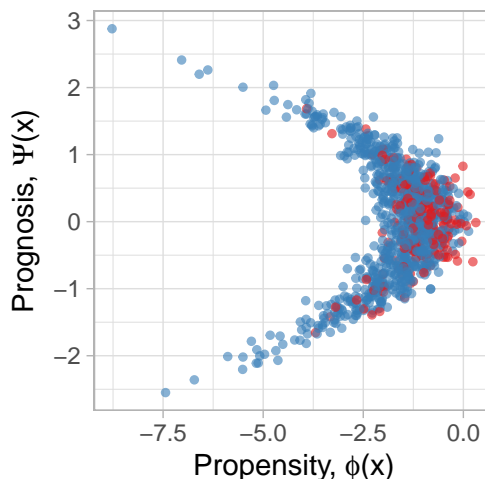
Figure 1 shows example assignment-control plots (A-C) and propensity score density histograms (D-F) for three simulated observational data sets. Notably, the marginal distribution of the propensity score is identical across the three simulated scenarios, and the corresponding propensity score histograms (D-F) are qualitatively equivalent. However, panels A-C reveal that the three settings are in fact quite different. In setting A, treated individuals are no different from control individuals in terms of their prognosis. This means that even a naive comparison of treated and control subjects may give an unbiased treatment effect estimate under suitable conditions. In settings B and C, treatment and control outcomes should not be directly compared, because the results would give a biased estimate of treatment effect. Moreover, the direction of the correlation between propensity and prognosis suggests the direction of the bias from a naive comparison: If the individuals with the *worst* prognosis are the *least* likely to be treated (Figure 1B), we are more likely to *overestimate* the effectiveness of the treatment in producing a more positive outcome. If the individuals with the *worst* prognosis are the *most* likely to be treated (Figure 1C), we are liable to err in the *opposite* direction. Future work might consider whether this correlation indicates a tendency toward bias not only in naive comparisons, but for adjustment or subclassification approaches in which the score models are imperfect or matchings of treated and control individuals are not exact.

2.2 Figure 2: Other Notable Scenarios

A



B



pdf
2

2.3 Figure 3: Assignment-Control Plots and Matching

pdf
2

Assignment-Control Plots can also be a useful diagnostic tool for matching studies. The four panels in figure 2 show the assignment-control visualizations of four different 1-to-1 matching schemes on the same data set: Mahalanobis distance (A), propensity score (B), Mahalanobis distance with a propensity score caliper (C), and Mahalanobis distance with both a propensity score caliper and a prognostic score caliper (D).

In Mahalanobis distance matching, all covariates are weighted equally in a statistical sense. When there is an abundance of uninformative covariates (i.e. those which are neither associated with the outcome nor the treatment assignment.), Mahalanobis distance matching can select matches that may actually be quite distant in the assignment-control space (Aikens, Greaves, and Baiocchi 2020). On the other hand, propensity score matching optimizes directly for matches which are nearby in terms of the variation associated with the treatment (the “assignment” axis), but it is entirely agnostic to variation associated with the outcome. This can result in high variance in estimated treatment effect (King and Nielsen 2016). Finally, the two caliper methods impose constraints on the matching option to ensure that matches are close in terms of propensity score (C) or both propensity and prognostic score (D). This visualization illustrates the potential of these methods for minimizing bias (stemming from poor propensity score balance) and variance (stemming from poor prognostic score balance) in the treatment effect.

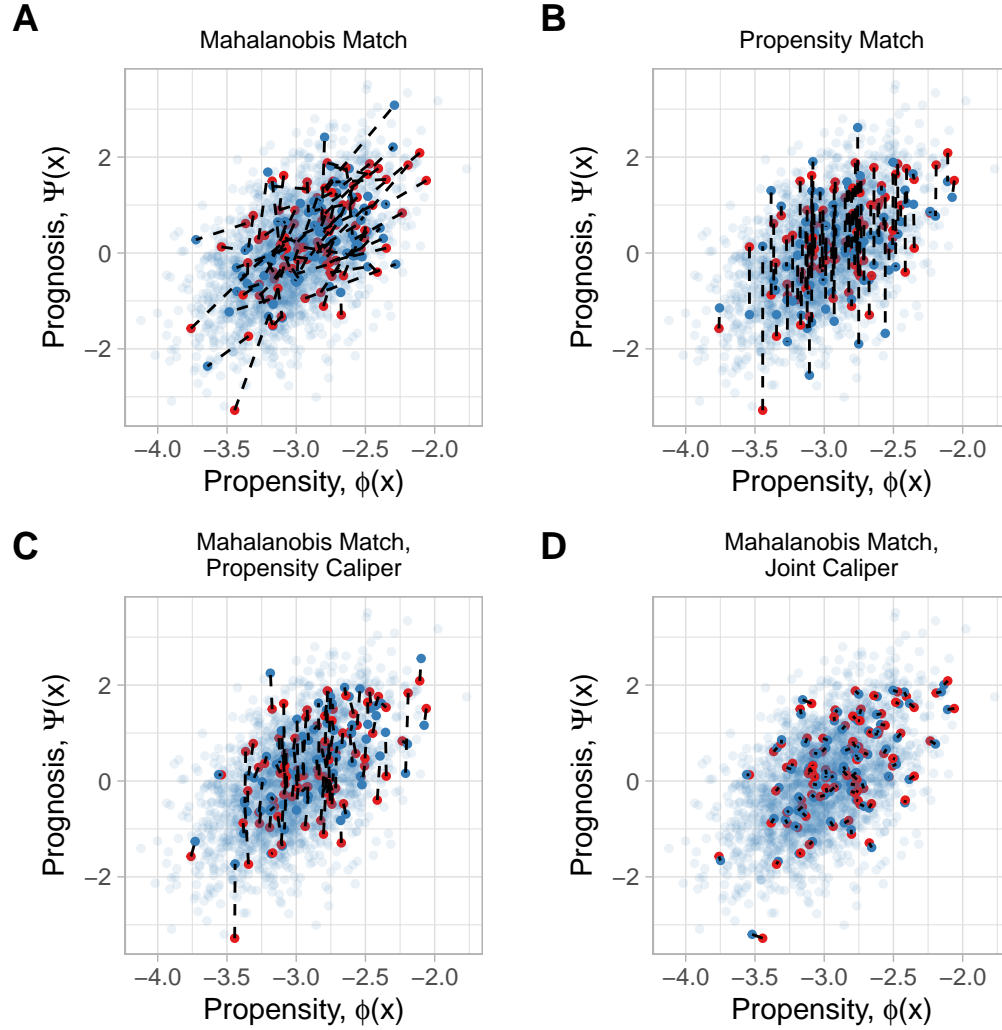
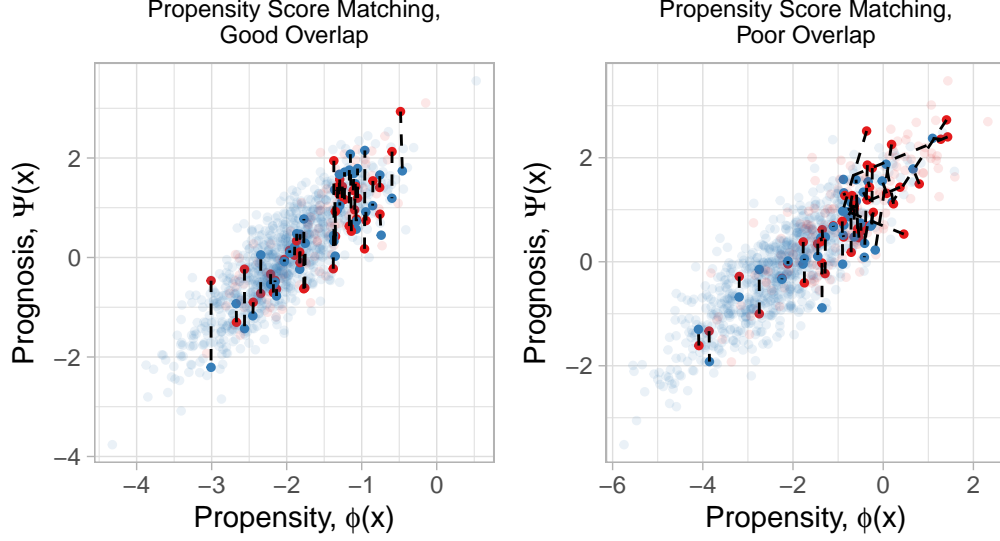


Figure 2: Assignment control plots depicting four different 1-to-1 matching schemes. Red points represent treated observations, blue points represent control. Dotted lines connect matched individuals. (A) Mahalanobis distance matching, (B), propensity score matching, (C) Mahalanobis distance matching with a propensity score caliper, (D) Mahalanobis distance matching with propensity and prognostic score calipers.



```
## pdf
## 2
```

2.4 Figure 5: Assignment-Control Plots and Unmeasured Confounding

There is a wide and increasing variety of methods which use some articulation of propensity and prognosis to estimate a treatment effect. However, the theorems underlying the use of the propensity score and the theorems underlying the use of the prognostic score both depend on the absence of unmeasured confounding. Violations of this assumption have ramifications for a wide variety of causal inference approaches, and assignment-control plots may be similarly misleading when unmeasured confounding is at play.

Figure 5 illustrates the behavior of assignment control plots in a scenario with unobserved confounding. We add to our data-generating set-up an unobserved confounder, U , such that:

$$\begin{aligned}\phi(X_i) &= c_1 X_{i1} + \eta U - c_0, \\ \Psi(X_i) &= \rho X_{i1} + \sqrt{(1 - \rho^2)} X_{i2} + \eta U.\end{aligned}$$

Suppose we somehow ascertained exactly the correct relationships between the two score models and the *observed* covariates, so that our propensity and prognostic models are precisely $\hat{\phi}(X_i) = c_1 X_{i1} - c_0$ and $\hat{\Psi}(X_i) = \rho X_{i1} + \sqrt{(1 - \rho^2)} X_{i2}$, respectively. That is, the score models are exactly correct, except that they do not include the unobserved confounder. Figure 5 panels A-C depict the assignment control plots we might make and matchings we might produce using these score models, $\hat{\phi}$ and $\hat{\Psi}$. Since both the assignment-control plots and the matchings use the score models $\hat{\phi}$ and $\hat{\Psi}$ that fail to capture unobserved confounding, our propensity matches appear quite close in $\hat{\phi}$ (Figure 5B) and our mahalanobis distance matchings with propensity and prognostic score calipers appear quite close in the assignment-control space defined by $\hat{\phi} \times \hat{\Psi}$ (Figure 5C).

However, panels D-F in Figure 5 show the same matches in the *true* assignment-control space, in which ϕ and Ψ are known to depend on the unobserved confounder, U . In each matching, pairs tend to differ from each other due to baseline variations in the unobserved confounder which were not accounted for in the matching process. The contrast between Figures 5C and 5F most clearly illustrate how failing to account for U results in systematic error: in the true assignment-control space, one matched individual in each pair tends to have both higher prognostic score and higher propensity score than its partner. Since this individual is more often the treated individual than the control individual, estimates of treatment effect based on this matching will tend to be biased, since U induces systematic differences between paired individuals, even after matching.

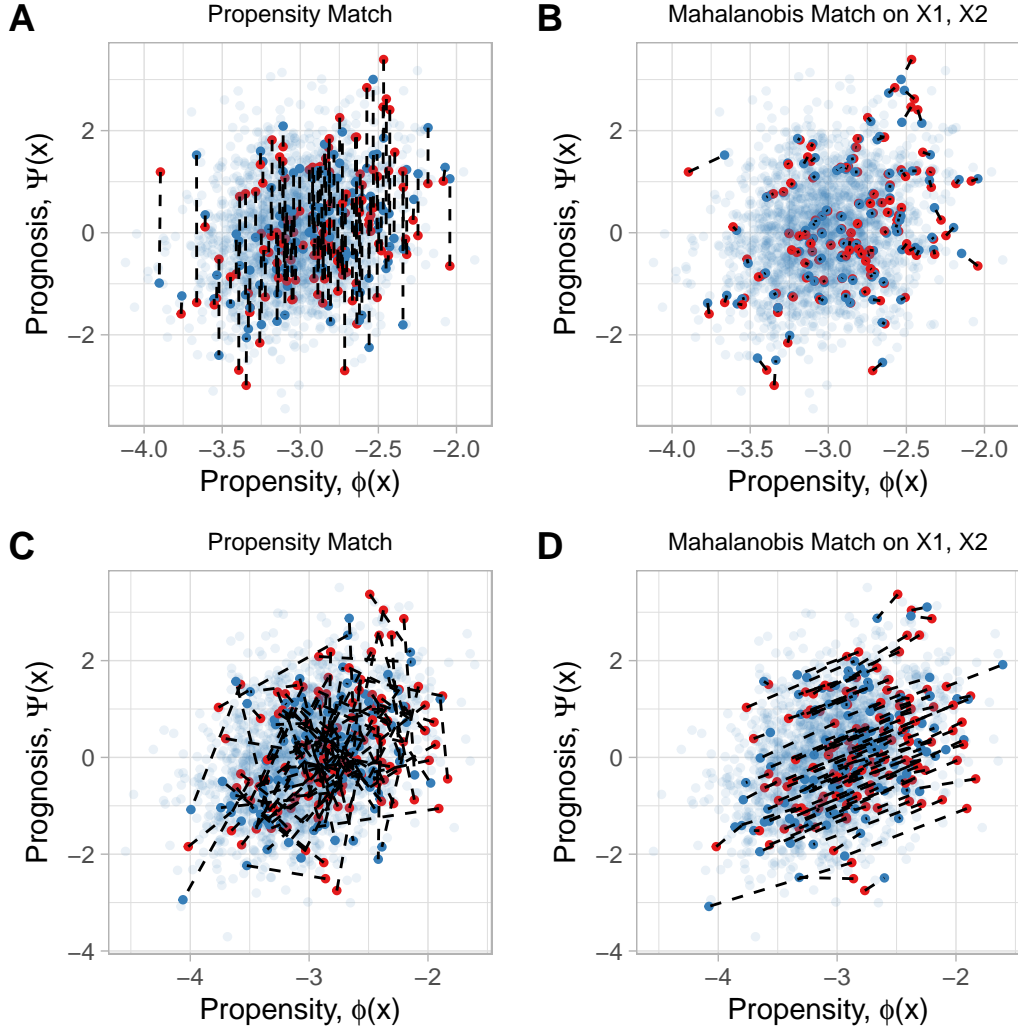


Figure 3: Assignment-control plots for three matching schemes on a dataset with unobserved confounding. A-C depict the assignment-control space as ascertained without knowledge of the unobserved confounder. D-F depict the true assignment-control space and the true match distances.

```
## pdf
## 2
```

2.5 Figure 6: Randomization-Assignment-Control Plots

We take the same set-up as before:

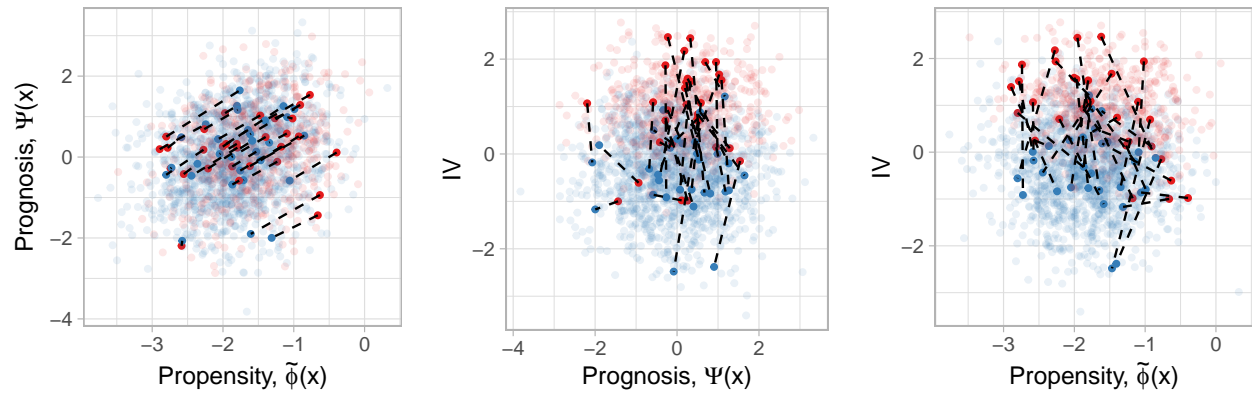
$$\begin{aligned}X_i &\sim \text{Normal}(0, I_{10}) \\T_i &\sim \text{Bernoulli}\left(\frac{1}{1 + \exp(\phi(X_i))}\right) \\Y_i(0) &= \Psi(X_i) + \epsilon_i \\ \epsilon_i &\sim N(0, 1)\end{aligned}$$

where $\phi(X)$ and $\Psi(X)$ represent the true propensity and prognostic score functions, given by

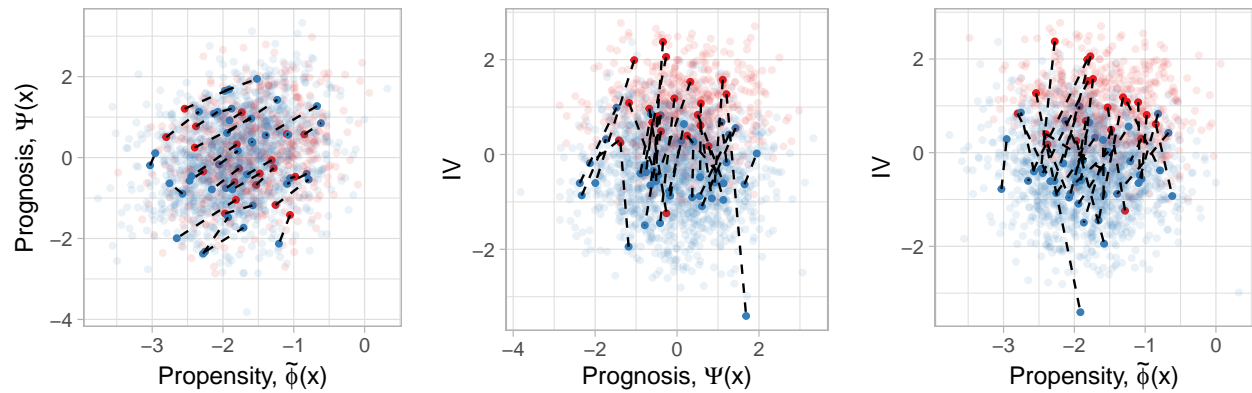
$$\begin{aligned}\phi(X_i, Z_i) &= c_1 X_{i1} + c_2 Z_i + \eta U - c_0, \\ \Psi(X_i) &= \rho X_{i1} + \sqrt{(1 - \rho^2)} X_{i2} + \eta U.\end{aligned}$$

Where c_2 , c_1 , c_0 , η and ρ are constants. Note that U represents an unmeasured confounder with strength η . As before, Z is an instrumental variable in that it is unconfounded and has no effect on the outcome except via the treatment. For simplicity, we'll make Z completely uncorrelated with the other covariates, so it is a “pure” IV. See `iv_with_correlated_covariates.Rmd` for a set-up that changes this.

A Mahalanobis Match on X1 and X2, Randomization–Assignment–Control–plots



B Nearfar Match, Randomization–Control–Assignment–plots



```
## pdf
## 2
```

2.6 Figures 7-8: Applied example

See `/analyses/CABG/CABG_women.Rmd`

References

Aikens, Rachael C, Dylan Greaves, and Michael Baiocchi. 2020. “A Pilot Design for Observational Studies: Using Abundant Data Thoughtfully.” *Statistics in Medicine*. Wiley Online Library.

King, Gary, and Richard Nielsen. 2016. “Why Propensity Scores Should Not Be Used for Matching.” *Copy at [Http://J. Mp/1sexgVw](http://J.Mp/1sexgVw) Download Citation BibTex Tagged XML Download Paper 378.*