MidtermReport

Lee Che Yuet Isaac

3/31/2023

# Progress so far

Preparation for the datasets are almost done, then can be moved on to the matching phase. There is a minor problem regarding including states as a covariate for matching, which is still considering for a solution.

# Initialization

rm(list=ls())  
gc()

# Vaccine incentives

## Loading the dataset as tibble

# load dataset  
vaccine\_incentives <- read\_delim("analysis\_ready.csv",  
delim = ";", escape\_double = FALSE, trim\_ws = TRUE) %>% # need to change the folder path  
 as\_tibble()  
  
#glimpse  
head(vaccine\_incentives)

## # A tibble: 6 × 139  
## ResponseID state file StartDate EndDate Progress Duration (in seconds…¹  
## <dbl> <chr> <dbl> <chr> <chr> <dbl> <dbl>  
## 1 1 Alabama 5 11.07.2021 1… 11.07.… 100 135  
## 2 2 Alabama 5 06.07.2021 0… 06.07.… 100 112  
## 3 3 Alabama 5 03.07.2021 1… 03.07.… 100 147  
## 4 4 Alabama 5 03.07.2021 1… 03.07.… 100 183  
## 5 5 Alabama 2 23.06.2021 0… 23.06.… 100 117  
## 6 6 Alabama 5 04.07.2021 0… 04.07.… 100 141  
## # ℹ abbreviated name: ¹​`Duration (in seconds)`  
## # ℹ 132 more variables: Finished <chr>, RecordedDate <chr>,  
## # RecipientLastName <lgl>, DistributionChannel <chr>, UserLanguage <chr>,  
## # Q\_RecaptchaScore <dbl>, Q\_RecaptchaAssessmentName <lgl>,  
## # Q\_RelevantIDDuplicate <lgl>, Q\_RelevantIDDuplicateScore <dbl>,  
## # Q\_RelevantIDFraudScore <dbl>, Q\_RelevantIDLastStartDate <chr>, Q1.2 <chr>,  
## # Q1.3 <chr>, Q84 <chr>, Q87 <chr>, Q135 <chr>, Q136 <chr>, age <dbl>, …

count(vaccine\_incentives, Treatment)

## # A tibble: 3 × 2  
## Treatment n  
## <chr> <int>  
## 1 treatment1 516  
## 2 treatment2 562  
## 3 treatment3 531

count(vaccine\_incentives, Treatment2)

## # A tibble: 3 × 2  
## Treatment2 n  
## <chr> <int>  
## 1 CDC Health Information 516  
## 2 Cash Voucher Incentive 531  
## 3 Lottery Incentive 562

Control: treatment1(CDC Health Information)  
Treatment version 1: treatment2(Lottery Incentive)  
Treatment version 2: treatment3(Cash Voucher Incentive)

## Data cleaning

vi\_cleaned <- vaccine\_incentives %>%  
 mutate(  
 Treatment\_012 = case\_when(Treatment == 'treatment1' ~ 0,  
 Treatment == 'treatment2' ~ 1,  
 Treatment == 'treatment3' ~ 2),  
 Female\_10 = case\_when(gender == 'Female' ~ 1,  
 gender == 'Male' ~ 0,  
 gender == 'Other' ~ NA\_real\_),  
 age\_young\_10 = if\_else(Age\_cat == 'Young', 1, 0),  
 race\_white\_10 = if\_else(Race == 'White' , 1, 0),  
 race\_black\_10 = if\_else(Race == 'Black' , 1, 0),  
 race\_other\_10 = if\_else(Race == 'Other' , 1, 0),  
 education\_low\_10 = if\_else(Education == 'Low' , 1, 0),  
 education\_medium\_10 = if\_else(Education == 'Medium' , 1, 0),  
 education\_high\_10 = if\_else(Education == 'High' , 1,0),  
 pool\_CloudResearch\_10 = if\_else(pool == 'CloudResearch' , 1, 0),  
 pool\_Facebook\_10 = if\_else(pool == 'Facebook' , 1, 0),  
 pool\_Lucid\_10 = if\_else(pool == 'Lucid' , 1, 0),  
 outcome = if\_else(clicked == 1, 1, 0),  
 treated = if\_else(Treatment\_012 > 0, 1, 0)) %>%  
 dplyr::select(Treatment\_012,  
 Female\_10,  
 age\_young\_10,  
 race\_white\_10,  
 race\_black\_10,  
 race\_other\_10,  
 education\_low\_10,  
 education\_medium\_10,  
 education\_high\_10,  
 pool\_CloudResearch\_10,  
 pool\_Facebook\_10,  
 pool\_Lucid\_10,  
 Trumppercent,  
 Trumphi,  
 age,  
 outcome,  
 treated)  
  
head(vi\_cleaned)

## # A tibble: 6 × 17  
## Treatment\_012 Female\_10 age\_young\_10 race\_white\_10 race\_black\_10 race\_other\_10  
## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 1 0 0 1 0 0  
## 2 1 0 1 1 0 0  
## 3 1 0 0 1 0 0  
## 4 1 1 1 0 1 0  
## 5 2 0 1 1 0 0  
## 6 2 1 0 1 0 0  
## # ℹ 11 more variables: education\_low\_10 <dbl>, education\_medium\_10 <dbl>,  
## # education\_high\_10 <dbl>, pool\_CloudResearch\_10 <dbl>,  
## # pool\_Facebook\_10 <dbl>, pool\_Lucid\_10 <dbl>, Trumppercent <dbl>,  
## # Trumphi <dbl>, age <dbl>, outcome <dbl>, treated <dbl>

## Summary statistics of whole dataset

vi\_cleaned %>%  
 mutate(  
 Treatment\_012 = factor(Treatment\_012),  
 Female\_10 = factor(Female\_10),  
 age\_young\_10 = factor(age\_young\_10),  
 race\_white\_10 = factor(race\_white\_10),  
 race\_black\_10 = factor(race\_black\_10),  
 race\_other\_10 = factor(race\_other\_10),  
 education\_low\_10 = factor(education\_low\_10),  
 education\_medium\_10 = factor(education\_medium\_10),  
 education\_high\_10 = factor(education\_high\_10),  
 pool\_CloudResearch\_10 = factor(pool\_CloudResearch\_10),  
 pool\_Facebook\_10 = factor(pool\_Facebook\_10),  
 pool\_Lucid\_10 = factor(pool\_Lucid\_10),  
 Trumppercent,  
 Trumphi = factor(Trumphi),  
 age,  
 outcome = factor(outcome),  
 treated = factor(treated)) %>%  
 summary()

## Treatment\_012 Female\_10 age\_young\_10 race\_white\_10 race\_black\_10  
## 0:516 0 : 540 0:680 0: 408 0:1326   
## 1:562 1 :1067 1:929 1:1201 1: 283   
## 2:531 NA's: 2   
##   
##   
##   
## race\_other\_10 education\_low\_10 education\_medium\_10 education\_high\_10  
## 0:1484 0 :949 0 :906 0 :1341   
## 1: 125 1 :649 1 :692 1 : 257   
## NA's: 11 NA's: 11 NA's: 11   
##   
##   
##   
## pool\_CloudResearch\_10 pool\_Facebook\_10 pool\_Lucid\_10 Trumppercent   
## 0: 188 0:1550 0:1480 Min. :3.190e+13   
## 1:1421 1: 59 1: 129 1st Qu.:3.920e+14   
## Median :4.910e+14   
## Mean :4.427e+14   
## 3rd Qu.:5.330e+14   
## Max. :6.950e+14   
## Trumphi age outcome treated   
## 0:778 Min. :18.00 0:1338 0: 516   
## 1:831 1st Qu.:29.00 1: 271 1:1093   
## Median :37.00   
## Mean :38.99   
## 3rd Qu.:47.00   
## Max. :90.00

## Control vs Treatment 1 & 2 (vi)

### Experimental benchmark

Creating a basic and preliminary estimate of the treatment effect.

model\_vi\_1 <- lm(outcome ~ treated, data = vi\_cleaned)  
summary(model\_vi\_1)

##   
## Call:  
## lm(formula = outcome ~ treated, data = vi\_cleaned)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.1738 -0.1738 -0.1738 -0.1570 0.8430   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.15698 0.01648 9.524 <2e-16 \*\*\*  
## treated 0.01686 0.02000 0.843 0.399   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3744 on 1607 degrees of freedom  
## Multiple R-squared: 0.000442, Adjusted R-squared: -0.00018   
## F-statistic: 0.7106 on 1 and 1607 DF, p-value: 0.3994

Try adding other covariates into the regression and se

model\_vi\_2 <- lm(outcome ~ treated +   
 Female\_10 +  
 race\_white\_10 +  
 race\_black\_10 +  
 education\_low\_10 +  
 education\_medium\_10 +  
 pool\_CloudResearch\_10 +  
 pool\_Facebook\_10 +  
 Trumppercent +  
 age, data = vi\_cleaned)  
summary(model\_vi\_2)

##   
## Call:  
## lm(formula = outcome ~ treated + Female\_10 + race\_white\_10 +   
## race\_black\_10 + education\_low\_10 + education\_medium\_10 +   
## pool\_CloudResearch\_10 + pool\_Facebook\_10 + Trumppercent +   
## age, data = vi\_cleaned)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.55213 -0.18165 -0.14951 -0.09802 0.96246   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -9.750e-02 6.503e-02 -1.499 0.134026   
## treated 1.529e-02 1.980e-02 0.773 0.439887   
## Female\_10 8.837e-03 1.977e-02 0.447 0.655011   
## race\_white\_10 -1.239e-02 3.544e-02 -0.350 0.726665   
## race\_black\_10 3.442e-02 4.002e-02 0.860 0.389882   
## education\_low\_10 4.384e-02 2.750e-02 1.594 0.111108   
## education\_medium\_10 2.531e-03 2.699e-02 0.094 0.925317   
## pool\_CloudResearch\_10 1.175e-01 3.460e-02 3.396 0.000701 \*\*\*  
## pool\_Facebook\_10 4.163e-01 5.875e-02 7.085 2.08e-12 \*\*\*  
## Trumppercent 6.092e-17 5.684e-17 1.072 0.283953   
## age 2.255e-03 7.073e-04 3.188 0.001462 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3681 on 1585 degrees of freedom  
## (13 observations deleted due to missingness)  
## Multiple R-squared: 0.03996, Adjusted R-squared: 0.0339   
## F-statistic: 6.597 on 10 and 1585 DF, p-value: 4.43e-10

### Matching preparation

vi\_tr <- which(vi\_cleaned$treated == 1)  
vi\_ct <- which(vi\_cleaned$treated == 0)  
vi\_ntr <- length(vi\_tr)  
vi\_ntr

## [1] 1093

vi\_nct <- length(vi\_ct)  
vi\_nct

## [1] 516

### Basic estimate

mean(na.omit(vi\_cleaned$outcome[vi\_tr])) - mean(na.omit(vi\_cleaned$outcome[vi\_ct]))

## [1] 0.01685674

### Covariate balance

vi\_covs1 <- vi\_cleaned %>%  
 dplyr::select(Female\_10,  
 race\_white\_10,  
 race\_black\_10,  
 race\_other\_10,  
 education\_low\_10,  
 education\_medium\_10,  
 education\_high\_10,  
 pool\_CloudResearch\_10,  
 pool\_Facebook\_10,  
 pool\_Lucid\_10,  
 Trumppercent,  
 age,  
 treated) %>%  
 na.omit()  
  
bal.tab(vi\_covs1, treat = vi\_covs1$treated)

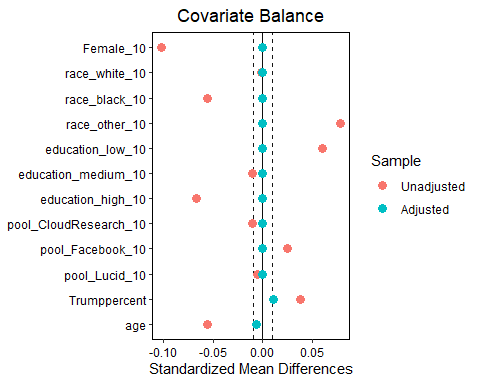
## Balance Measures  
## Type Diff.Un  
## Female\_10 Binary -0.0487  
## race\_white\_10 Binary -0.0006  
## race\_black\_10 Binary -0.0210  
## race\_other\_10 Binary 0.0216  
## education\_low\_10 Binary 0.0295  
## education\_medium\_10 Binary -0.0054  
## education\_high\_10 Binary -0.0241  
## pool\_CloudResearch\_10 Binary -0.0033  
## pool\_Facebook\_10 Binary 0.0047  
## pool\_Lucid\_10 Binary -0.0014  
## Trumppercent Contin. 0.0364  
## age Contin. -0.0544  
## treated Binary 1.0000  
##   
## Sample sizes  
## Control Treated  
## All 513 1083

### Coarsened Exact Matching

vi\_data1 <- na.omit(vi\_cleaned)  
  
vi\_X\_1 <- vi\_data1 %>%  
 dplyr::select(Female\_10,  
 race\_white\_10,  
 race\_black\_10,  
 race\_other\_10,  
 education\_low\_10,  
 education\_medium\_10,  
 education\_high\_10,  
 pool\_CloudResearch\_10,  
 pool\_Facebook\_10,  
 pool\_Lucid\_10,  
 Trumppercent,  
 age)  
  
vi\_m.out.cem <- matchit(treated ~ Female\_10 +  
 race\_white\_10 +  
 race\_black\_10 +  
 race\_other\_10 +  
 education\_low\_10 +  
 education\_medium\_10 +  
 education\_high\_10 +  
 pool\_CloudResearch\_10 +  
 pool\_Facebook\_10 +  
 pool\_Lucid\_10 +  
 Trumppercent +  
 age,  
 data = vi\_data1,  
 method = "cem")  
  
vi\_bal\_tab\_cem <- bal.tab(vi\_m.out.cem,  
 thresholds = c(m = 0.01),  
 un = TRUE)  
vi\_bal\_tab\_cem

## Balance Measures  
## Type Diff.Un Diff.Adj M.Threshold  
## Female\_10 Binary -0.0487 0.0000 Balanced, <0.01  
## race\_white\_10 Binary -0.0006 -0.0000 Balanced, <0.01  
## race\_black\_10 Binary -0.0210 -0.0000 Balanced, <0.01  
## race\_other\_10 Binary 0.0216 0.0000 Balanced, <0.01  
## education\_low\_10 Binary 0.0295 0.0000 Balanced, <0.01  
## education\_medium\_10 Binary -0.0054 0.0000 Balanced, <0.01  
## education\_high\_10 Binary -0.0241 -0.0000 Balanced, <0.01  
## pool\_CloudResearch\_10 Binary -0.0033 -0.0000 Balanced, <0.01  
## pool\_Facebook\_10 Binary 0.0047 0.0000 Balanced, <0.01  
## pool\_Lucid\_10 Binary -0.0014 0.0000 Balanced, <0.01  
## Trumppercent Contin. 0.0378 0.0108 Not Balanced, >0.01  
## age Contin. -0.0561 -0.0068 Balanced, <0.01  
##   
## Balance tally for mean differences  
## count  
## Balanced, <0.01 11  
## Not Balanced, >0.01 1  
##   
## Variable with the greatest mean difference  
## Variable Diff.Adj M.Threshold  
## Trumppercent 0.0108 Not Balanced, >0.01  
##   
## Sample sizes  
## Control Treated  
## All 513. 1083  
## Matched (ESS) 232.55 616  
## Matched (Unweighted) 371. 616  
## Unmatched 142. 467

vi\_lp.cem <- love.plot(  
 vi\_m.out.cem,  
 threshold = .01,  
 binary = 'std',  
 treat = vi\_data1$treated,  
 covs = vi\_X\_1)  
  
vi\_lp.cem



#### ATT by CEM

vi\_m.data.cem <- match.data(vi\_m.out.cem)  
head(vi\_m.data.cem)

## # A tibble: 6 × 19  
## Treatment\_012 Female\_10 age\_young\_10 race\_white\_10 race\_black\_10 race\_other\_10  
## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 1 0 0 1 0 0  
## 2 1 0 1 1 0 0  
## 3 1 0 0 1 0 0  
## 4 2 1 0 1 0 0  
## 5 1 1 0 1 0 0  
## 6 0 0 1 0 1 0  
## # ℹ 13 more variables: education\_low\_10 <dbl>, education\_medium\_10 <dbl>,  
## # education\_high\_10 <dbl>, pool\_CloudResearch\_10 <dbl>,  
## # pool\_Facebook\_10 <dbl>, pool\_Lucid\_10 <dbl>, Trumppercent <dbl>,  
## # Trumphi <dbl>, age <dbl>, outcome <dbl>, treated <dbl>, weights <dbl>,  
## # subclass <fct>

#### CEM regression

vi\_lm.cem1 <- lm(outcome ~ treated, data = vi\_m.data.cem)  
summary(vi\_lm.cem1)

##   
## Call:  
## lm(formula = outcome ~ treated, data = vi\_m.data.cem)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.1769 -0.1769 -0.1769 -0.1509 0.8491   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.15094 0.01938 7.788 1.71e-14 \*\*\*  
## treated 0.02600 0.02453 1.060 0.289   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3733 on 985 degrees of freedom  
## Multiple R-squared: 0.001139, Adjusted R-squared: 0.0001254   
## F-statistic: 1.124 on 1 and 985 DF, p-value: 0.2894

vi\_lm.cem2 <- lm(outcome ~ treated +   
 Female\_10 +  
 race\_white\_10 +  
 race\_black\_10 +  
 education\_low\_10 +  
 education\_medium\_10 +  
 pool\_CloudResearch\_10 +  
 pool\_Facebook\_10 +  
 Trumppercent +  
 age, data = vi\_m.data.cem)  
summary(vi\_lm.cem2)

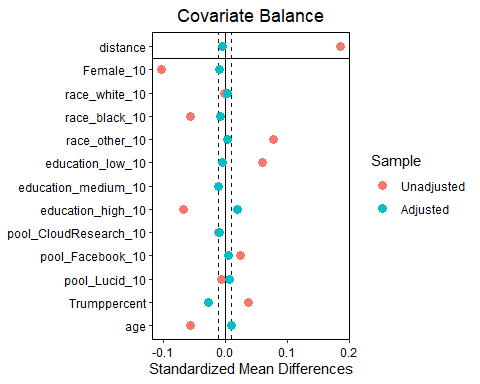
##   
## Call:  
## lm(formula = outcome ~ treated + Female\_10 + race\_white\_10 +   
## race\_black\_10 + education\_low\_10 + education\_medium\_10 +   
## pool\_CloudResearch\_10 + pool\_Facebook\_10 + Trumppercent +   
## age, data = vi\_m.data.cem)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.4204 -0.1907 -0.1506 -0.1017 0.9365   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.578e-02 1.366e-01 0.262 0.79338   
## treated 2.577e-02 2.456e-02 1.049 0.29422   
## Female\_10 2.039e-03 2.652e-02 0.077 0.93872   
## race\_white\_10 -1.285e-01 1.007e-01 -1.277 0.20194   
## race\_black\_10 -4.730e-02 1.055e-01 -0.448 0.65413   
## education\_low\_10 7.422e-02 3.922e-02 1.892 0.05878 .   
## education\_medium\_10 2.013e-02 3.897e-02 0.517 0.60552   
## pool\_CloudResearch\_10 6.637e-02 6.961e-02 0.953 0.34066   
## pool\_Facebook\_10 1.325e-01 1.981e-01 0.669 0.50363   
## Trumppercent 3.052e-17 7.694e-17 0.397 0.69165   
## age 2.875e-03 1.045e-03 2.753 0.00602 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3717 on 976 degrees of freedom  
## Multiple R-squared: 0.01862, Adjusted R-squared: 0.008564   
## F-statistic: 1.852 on 10 and 976 DF, p-value: 0.04829

### Genetic matching

set.seed(1)  
vi\_m.out.gen <- matchit(treated ~ Female\_10 +  
 race\_white\_10 +  
 race\_black\_10 +  
 race\_other\_10 +  
 education\_low\_10 +  
 education\_medium\_10 +  
 education\_high\_10 +  
 pool\_CloudResearch\_10 +  
 pool\_Facebook\_10 +  
 pool\_Lucid\_10 +  
 Trumppercent +  
 age,   
 data = vi\_data1,   
 replace = TRUE,   
 method = "genetic")  
vi\_bal\_tab\_gen <- bal.tab(vi\_m.out.gen, thresholds = c(m = 0.01), un = TRUE)  
vi\_bal\_tab\_gen

## Balance Measures  
## Type Diff.Un Diff.Adj M.Threshold  
## distance Distance 0.1857 -0.0042 Balanced, <0.01  
## Female\_10 Binary -0.0487 -0.0046 Balanced, <0.01  
## race\_white\_10 Binary -0.0006 0.0018 Balanced, <0.01  
## race\_black\_10 Binary -0.0210 -0.0028 Balanced, <0.01  
## race\_other\_10 Binary 0.0216 0.0009 Balanced, <0.01  
## education\_low\_10 Binary 0.0295 -0.0018 Balanced, <0.01  
## education\_medium\_10 Binary -0.0054 -0.0055 Balanced, <0.01  
## education\_high\_10 Binary -0.0241 0.0074 Balanced, <0.01  
## pool\_CloudResearch\_10 Binary -0.0033 -0.0028 Balanced, <0.01  
## pool\_Facebook\_10 Binary 0.0047 0.0009 Balanced, <0.01  
## pool\_Lucid\_10 Binary -0.0014 0.0018 Balanced, <0.01  
## Trumppercent Contin. 0.0378 -0.0263 Not Balanced, >0.01  
## age Contin. -0.0561 0.0100 Not Balanced, >0.01  
##   
## Balance tally for mean differences  
## count  
## Balanced, <0.01 11  
## Not Balanced, >0.01 2  
##   
## Variable with the greatest mean difference  
## Variable Diff.Adj M.Threshold  
## Trumppercent -0.0263 Not Balanced, >0.01  
##   
## Sample sizes  
## Control Treated  
## All 513. 1083  
## Matched (ESS) 259.32 1083  
## Matched (Unweighted) 396. 1083  
## Unmatched 117. 0

vi\_lp.gen <- love.plot(  
 vi\_m.out.gen,  
 threshold = .01,  
 binary = 'std',  
 treat = vi\_data1$treated,  
 covs = vi\_X\_1)  
  
vi\_lp.gen



#### ATT by Gen

vi\_m.data.gen <- match.data(vi\_m.out.gen)  
head(vi\_m.data.gen)

## # A tibble: 6 × 19  
## Treatment\_012 Female\_10 age\_young\_10 race\_white\_10 race\_black\_10 race\_other\_10  
## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 1 0 0 1 0 0  
## 2 1 0 1 1 0 0  
## 3 1 0 0 1 0 0  
## 4 1 1 1 0 1 0  
## 5 2 0 1 1 0 0  
## 6 2 1 0 1 0 0  
## # ℹ 13 more variables: education\_low\_10 <dbl>, education\_medium\_10 <dbl>,  
## # education\_high\_10 <dbl>, pool\_CloudResearch\_10 <dbl>,  
## # pool\_Facebook\_10 <dbl>, pool\_Lucid\_10 <dbl>, Trumppercent <dbl>,  
## # Trumphi <dbl>, age <dbl>, outcome <dbl>, treated <dbl>, distance <dbl>,  
## # weights <dbl>

#### Gen regression

vi\_lm.gen1 <- lm(outcome ~ treated, data = vi\_m.data.gen)  
summary(vi\_lm.gen1)

##   
## Call:  
## lm(formula = outcome ~ treated, data = vi\_m.data.gen)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.1736 -0.1736 -0.1736 -0.1490 0.8510   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.14899 0.01875 7.947 3.76e-15 \*\*\*  
## treated 0.02460 0.02191 1.123 0.262   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3731 on 1477 degrees of freedom  
## Multiple R-squared: 0.000853, Adjusted R-squared: 0.0001765   
## F-statistic: 1.261 on 1 and 1477 DF, p-value: 0.2616

vi\_lm.gen2 <- lm(outcome ~ treated +   
 Female\_10 +  
 race\_white\_10 +  
 race\_black\_10 +  
 education\_low\_10 +  
 education\_medium\_10 +  
 pool\_CloudResearch\_10 +  
 pool\_Facebook\_10 +  
 Trumppercent +  
 age, data = vi\_m.data.gen)  
summary(vi\_lm.gen2)

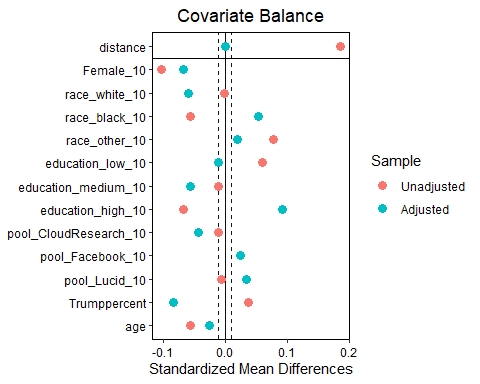
##   
## Call:  
## lm(formula = outcome ~ treated + Female\_10 + race\_white\_10 +   
## race\_black\_10 + education\_low\_10 + education\_medium\_10 +   
## pool\_CloudResearch\_10 + pool\_Facebook\_10 + Trumppercent +   
## age, data = vi\_m.data.gen)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.55072 -0.18241 -0.14826 -0.09406 0.96623   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.139e-01 6.748e-02 -1.687 0.091742 .   
## treated 2.217e-02 2.157e-02 1.028 0.304241   
## Female\_10 1.224e-02 2.037e-02 0.601 0.548151   
## race\_white\_10 -6.012e-03 3.623e-02 -0.166 0.868217   
## race\_black\_10 3.985e-02 4.127e-02 0.966 0.334415   
## education\_low\_10 3.759e-02 2.861e-02 1.314 0.189028   
## education\_medium\_10 -4.272e-03 2.816e-02 -0.152 0.879463   
## pool\_CloudResearch\_10 1.233e-01 3.549e-02 3.475 0.000526 \*\*\*  
## pool\_Facebook\_10 4.207e-01 6.051e-02 6.953 5.36e-12 \*\*\*  
## Trumppercent 6.840e-17 5.924e-17 1.155 0.248417   
## age 2.216e-03 7.437e-04 2.980 0.002933 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3666 on 1468 degrees of freedom  
## Multiple R-squared: 0.0411, Adjusted R-squared: 0.03457   
## F-statistic: 6.292 on 10 and 1468 DF, p-value: 1.666e-09

### Nearest neighbour matching

vi\_m.out.nnm <- matchit(treated ~ Female\_10 +  
 race\_white\_10 +  
 race\_black\_10 +  
 race\_other\_10 +  
 education\_low\_10 +  
 education\_medium\_10 +  
 education\_high\_10 +  
 pool\_CloudResearch\_10 +  
 pool\_Facebook\_10 +  
 pool\_Lucid\_10 +  
 Trumppercent +  
 age,   
 data = vi\_data1,   
 replace = TRUE,   
 method = "nearest")  
vi\_bal\_tab\_nnm <- bal.tab(vi\_m.out.nnm, thresholds = c(m = 0.01), un = TRUE)  
vi\_bal\_tab\_nnm

## Balance Measures  
## Type Diff.Un Diff.Adj M.Threshold  
## distance Distance 0.1857 0.0002 Balanced, <0.01  
## Female\_10 Binary -0.0487 -0.0323 Not Balanced, >0.01  
## race\_white\_10 Binary -0.0006 -0.0259 Not Balanced, >0.01  
## race\_black\_10 Binary -0.0210 0.0203 Not Balanced, >0.01  
## race\_other\_10 Binary 0.0216 0.0055 Balanced, <0.01  
## education\_low\_10 Binary 0.0295 -0.0055 Balanced, <0.01  
## education\_medium\_10 Binary -0.0054 -0.0277 Not Balanced, >0.01  
## education\_high\_10 Binary -0.0241 0.0332 Not Balanced, >0.01  
## pool\_CloudResearch\_10 Binary -0.0033 -0.0139 Not Balanced, >0.01  
## pool\_Facebook\_10 Binary 0.0047 0.0046 Balanced, <0.01  
## pool\_Lucid\_10 Binary -0.0014 0.0092 Balanced, <0.01  
## Trumppercent Contin. 0.0378 -0.0829 Not Balanced, >0.01  
## age Contin. -0.0561 -0.0244 Not Balanced, >0.01  
##   
## Balance tally for mean differences  
## count  
## Balanced, <0.01 5  
## Not Balanced, >0.01 8  
##   
## Variable with the greatest mean difference  
## Variable Diff.Adj M.Threshold  
## Trumppercent -0.0829 Not Balanced, >0.01  
##   
## Sample sizes  
## Control Treated  
## All 513. 1083  
## Matched (ESS) 270.56 1083  
## Matched (Unweighted) 395. 1083  
## Unmatched 118. 0

vi\_lp.nnm <- love.plot(  
 vi\_m.out.nnm,  
 threshold = .01,  
 binary = 'std',  
 treat = vi\_data1$treated,  
 covs = vi\_X\_1)  
  
vi\_lp.nnm



#### ATT by NNM

vi\_m.data.nnm <- match.data(vi\_m.out.nnm)  
head(vi\_m.data.nnm)

## # A tibble: 6 × 19  
## Treatment\_012 Female\_10 age\_young\_10 race\_white\_10 race\_black\_10 race\_other\_10  
## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 1 0 0 1 0 0  
## 2 1 0 1 1 0 0  
## 3 1 0 0 1 0 0  
## 4 1 1 1 0 1 0  
## 5 2 0 1 1 0 0  
## 6 2 1 0 1 0 0  
## # ℹ 13 more variables: education\_low\_10 <dbl>, education\_medium\_10 <dbl>,  
## # education\_high\_10 <dbl>, pool\_CloudResearch\_10 <dbl>,  
## # pool\_Facebook\_10 <dbl>, pool\_Lucid\_10 <dbl>, Trumppercent <dbl>,  
## # Trumphi <dbl>, age <dbl>, outcome <dbl>, treated <dbl>, distance <dbl>,  
## # weights <dbl>

#### NNM regression

vi\_lm.nnm1 <- lm(outcome ~ treated, data = vi\_m.data.nnm)  
summary(vi\_lm.nnm1)

##   
## Call:  
## lm(formula = outcome ~ treated, data = vi\_m.data.nnm)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.1736 -0.1736 -0.1736 -0.1544 0.8456   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.15443 0.01884 8.197 5.3e-16 \*\*\*  
## treated 0.01916 0.02201 0.871 0.384   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3744 on 1476 degrees of freedom  
## Multiple R-squared: 0.0005133, Adjusted R-squared: -0.0001639   
## F-statistic: 0.758 on 1 and 1476 DF, p-value: 0.3841

vi\_lm.nnm2 <- lm(outcome ~ treated +   
 Female\_10 +  
 race\_white\_10 +  
 race\_black\_10 +  
 education\_low\_10 +  
 education\_medium\_10 +  
 pool\_CloudResearch\_10 +  
 pool\_Facebook\_10 +  
 Trumppercent +  
 age, data = vi\_m.data.nnm)  
summary(vi\_lm.nnm2)

##   
## Call:  
## lm(formula = outcome ~ treated + Female\_10 + race\_white\_10 +   
## race\_black\_10 + education\_low\_10 + education\_medium\_10 +   
## pool\_CloudResearch\_10 + pool\_Facebook\_10 + Trumppercent +   
## age, data = vi\_m.data.nnm)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.55295 -0.18262 -0.14922 -0.09918 0.96534   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.116e-01 6.917e-02 -1.613 0.106908   
## treated 2.015e-02 2.169e-02 0.929 0.353102   
## Female\_10 1.362e-02 2.051e-02 0.664 0.506874   
## race\_white\_10 -5.592e-03 3.647e-02 -0.153 0.878167   
## race\_black\_10 3.951e-02 4.145e-02 0.953 0.340656   
## education\_low\_10 5.016e-02 2.904e-02 1.727 0.084332 .   
## education\_medium\_10 6.276e-03 2.858e-02 0.220 0.826249   
## pool\_CloudResearch\_10 1.205e-01 3.640e-02 3.311 0.000952 \*\*\*  
## pool\_Facebook\_10 4.179e-01 6.123e-02 6.826 1.28e-11 \*\*\*  
## Trumppercent 2.946e-17 6.019e-17 0.489 0.624577   
## age 2.441e-03 7.410e-04 3.294 0.001010 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3679 on 1467 degrees of freedom  
## Multiple R-squared: 0.041, Adjusted R-squared: 0.03447   
## F-statistic: 6.273 on 10 and 1467 DF, p-value: 1.812e-09

## Control vs Treatment 1 only (vi1)

### Experimental benchmark

Creating a basic and preliminary estimate of the treatment effect.

vi1\_cleaned <- vi\_cleaned %>%  
 filter(Treatment\_012 < 2)  
  
count(vi1\_cleaned, Treatment\_012)

## # A tibble: 2 × 2  
## Treatment\_012 n  
## <dbl> <int>  
## 1 0 516  
## 2 1 562

model\_vi1\_1 <- lm(outcome ~ treated, data = vi1\_cleaned)  
summary(model\_vi1\_1)

##   
## Call:  
## lm(formula = outcome ~ treated, data = vi1\_cleaned)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.1570 -0.1570 -0.1352 -0.1352 0.8648   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.15698 0.01554 10.104 <2e-16 \*\*\*  
## treated -0.02175 0.02152 -1.011 0.312   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3529 on 1076 degrees of freedom  
## Multiple R-squared: 0.0009483, Adjusted R-squared: 1.985e-05   
## F-statistic: 1.021 on 1 and 1076 DF, p-value: 0.3124

Try adding other covariates into the regression and see

model\_vi1\_2 <- lm(outcome ~ treated +   
 Female\_10 +  
 race\_white\_10 +  
 race\_black\_10 +  
 education\_low\_10 +  
 education\_medium\_10 +  
 pool\_CloudResearch\_10 +  
 pool\_Facebook\_10 +  
 Trumppercent +  
 age, data = vi1\_cleaned)  
summary(model\_vi1\_2)

##   
## Call:  
## lm(formula = outcome ~ treated + Female\_10 + race\_white\_10 +   
## race\_black\_10 + education\_low\_10 + education\_medium\_10 +   
## pool\_CloudResearch\_10 + pool\_Facebook\_10 + Trumppercent +   
## age, data = vi1\_cleaned)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.54454 -0.15580 -0.12208 -0.08314 0.98235   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -3.836e-02 7.402e-02 -0.518 0.6043   
## treated -2.673e-02 2.131e-02 -1.254 0.2101   
## Female\_10 -9.275e-03 2.288e-02 -0.405 0.6853   
## race\_white\_10 -2.867e-02 4.045e-02 -0.709 0.4785   
## race\_black\_10 3.238e-02 4.532e-02 0.714 0.4751   
## education\_low\_10 4.844e-02 3.091e-02 1.567 0.1174   
## education\_medium\_10 1.441e-02 3.033e-02 0.475 0.6348   
## pool\_CloudResearch\_10 8.252e-02 4.000e-02 2.063 0.0394 \*   
## pool\_Facebook\_10 4.173e-01 6.591e-02 6.332 3.58e-10 \*\*\*  
## Trumppercent 5.277e-17 6.506e-17 0.811 0.4174   
## age 2.095e-03 8.161e-04 2.568 0.0104 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3459 on 1056 degrees of freedom  
## (11 observations deleted due to missingness)  
## Multiple R-squared: 0.05126, Adjusted R-squared: 0.04227   
## F-statistic: 5.705 on 10 and 1056 DF, p-value: 2.265e-08

### Matching preparation

vi1\_tr <- which(vi1\_cleaned$Treatment\_012 == 1)  
vi1\_ct <- which(vi1\_cleaned$Treatment\_012 == 0)  
vi1\_ntr <- length(vi1\_tr)  
vi1\_ntr

## [1] 562

vi1\_nct <- length(vi1\_ct)  
vi1\_nct

## [1] 516

### Basic estimate

mean(na.omit(vi1\_cleaned$outcome[vi1\_tr])) - mean(na.omit(vi1\_cleaned$outcome[vi1\_ct]))

## [1] -0.02174543

### Covariate balance

vi1\_covs1 <- vi1\_cleaned %>%  
 dplyr::select(Female\_10,  
 race\_white\_10,  
 race\_black\_10,  
 race\_other\_10,  
 education\_low\_10,  
 education\_medium\_10,  
 education\_high\_10,  
 pool\_CloudResearch\_10,  
 pool\_Facebook\_10,  
 pool\_Lucid\_10,  
 Trumppercent,  
 age,  
 treated) %>%  
 na.omit()  
  
bal.tab(vi1\_covs1, treat = vi1\_covs1$treated)

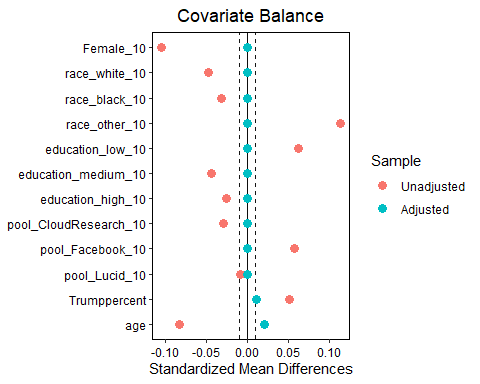
## Balance Measures  
## Type Diff.Un  
## Female\_10 Binary -0.0498  
## race\_white\_10 Binary -0.0211  
## race\_black\_10 Binary -0.0122  
## race\_other\_10 Binary 0.0333  
## education\_low\_10 Binary 0.0310  
## education\_medium\_10 Binary -0.0215  
## education\_high\_10 Binary -0.0095  
## pool\_CloudResearch\_10 Binary -0.0097  
## pool\_Facebook\_10 Binary 0.0120  
## pool\_Lucid\_10 Binary -0.0023  
## Trumppercent Contin. 0.0488  
## age Contin. -0.0800  
## treated Binary 1.0000  
##   
## Sample sizes  
## Control Treated  
## All 513 554

### CEM

vi1\_data1 <- na.omit(vi1\_cleaned)  
  
vi1\_X\_1 <- vi1\_data1 %>%  
 dplyr::select(Female\_10,  
 race\_white\_10,  
 race\_black\_10,  
 race\_other\_10,  
 education\_low\_10,  
 education\_medium\_10,  
 education\_high\_10,  
 pool\_CloudResearch\_10,  
 pool\_Facebook\_10,  
 pool\_Lucid\_10,  
 Trumppercent,  
 age)  
  
vi1\_m.out.cem <- matchit(treated ~ Female\_10 +  
 race\_white\_10 +  
 race\_black\_10 +  
 race\_other\_10 +  
 education\_low\_10 +  
 education\_medium\_10 +  
 education\_high\_10 +  
 pool\_CloudResearch\_10 +  
 pool\_Facebook\_10 +  
 pool\_Lucid\_10 +  
 Trumppercent +  
 age,  
 data = vi1\_data1,  
 method = "cem")  
  
vi1\_bal\_tab\_cem <- bal.tab(vi1\_m.out.cem,  
 thresholds = c(m = 0.01),  
 un = TRUE)  
vi1\_bal\_tab\_cem

## Balance Measures  
## Type Diff.Un Diff.Adj M.Threshold  
## Female\_10 Binary -0.0498 0.0000 Balanced, <0.01  
## race\_white\_10 Binary -0.0211 0.0000 Balanced, <0.01  
## race\_black\_10 Binary -0.0122 0.0000 Balanced, <0.01  
## race\_other\_10 Binary 0.0333 0.0000 Balanced, <0.01  
## education\_low\_10 Binary 0.0310 0.0000 Balanced, <0.01  
## education\_medium\_10 Binary -0.0215 0.0000 Balanced, <0.01  
## education\_high\_10 Binary -0.0095 0.0000 Balanced, <0.01  
## pool\_CloudResearch\_10 Binary -0.0097 -0.0000 Balanced, <0.01  
## pool\_Facebook\_10 Binary 0.0120 0.0000 Balanced, <0.01  
## pool\_Lucid\_10 Binary -0.0023 0.0000 Balanced, <0.01  
## Trumppercent Contin. 0.0509 0.0114 Not Balanced, >0.01  
## age Contin. -0.0830 0.0204 Not Balanced, >0.01  
##   
## Balance tally for mean differences  
## count  
## Balanced, <0.01 10  
## Not Balanced, >0.01 2  
##   
## Variable with the greatest mean difference  
## Variable Diff.Adj M.Threshold  
## age 0.0204 Not Balanced, >0.01  
##   
## Sample sizes  
## Control Treated  
## All 513. 554  
## Matched (ESS) 178.38 279  
## Matched (Unweighted) 297. 279  
## Unmatched 216. 275

vi1\_lp.cem <- love.plot(  
 vi1\_m.out.cem,  
 threshold = .01,  
 binary = 'std',  
 treat = vi1\_data1$treated,  
 covs = vi1\_X\_1)  
  
vi1\_lp.cem



#### ATT by CEM

vi1\_m.data.cem <- match.data(vi1\_m.out.cem)  
head(vi1\_m.data.cem)

## # A tibble: 6 × 19  
## Treatment\_012 Female\_10 age\_young\_10 race\_white\_10 race\_black\_10 race\_other\_10  
## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 1 0 1 1 0 0  
## 2 1 0 0 1 0 0  
## 3 1 1 0 1 0 0  
## 4 0 0 1 0 1 0  
## 5 1 1 1 0 1 0  
## 6 0 1 1 0 1 0  
## # ℹ 13 more variables: education\_low\_10 <dbl>, education\_medium\_10 <dbl>,  
## # education\_high\_10 <dbl>, pool\_CloudResearch\_10 <dbl>,  
## # pool\_Facebook\_10 <dbl>, pool\_Lucid\_10 <dbl>, Trumppercent <dbl>,  
## # Trumphi <dbl>, age <dbl>, outcome <dbl>, treated <dbl>, weights <dbl>,  
## # subclass <fct>

#### CEM regression

vi1\_lm.cem1 <- lm(outcome ~ treated, data = vi1\_m.data.cem)  
summary(vi1\_lm.cem1)

##   
## Call:  
## lm(formula = outcome ~ treated, data = vi1\_m.data.cem)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.1717 -0.1717 -0.1326 -0.1326 0.8674   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.17172 0.02088 8.223 1.33e-15 \*\*\*  
## treated -0.03910 0.03000 -1.303 0.193   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3599 on 574 degrees of freedom  
## Multiple R-squared: 0.00295, Adjusted R-squared: 0.001213   
## F-statistic: 1.698 on 1 and 574 DF, p-value: 0.193

vi1\_lm.cem2 <- lm(outcome ~ treated +   
 Female\_10 +  
 race\_white\_10 +  
 race\_black\_10 +  
 education\_low\_10 +  
 education\_medium\_10 +  
 pool\_CloudResearch\_10 +  
 pool\_Facebook\_10 +  
 Trumppercent +  
 age, data = vi1\_m.data.cem)  
summary(vi1\_lm.cem2)

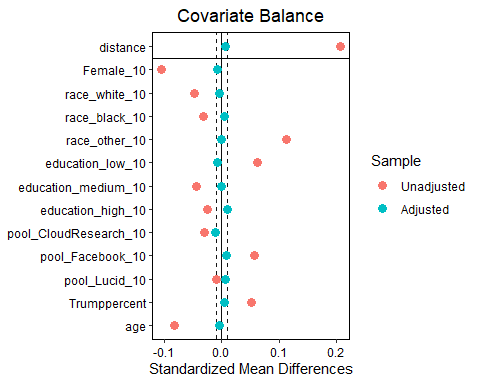
##   
## Call:  
## lm(formula = outcome ~ treated + Female\_10 + race\_white\_10 +   
## race\_black\_10 + education\_low\_10 + education\_medium\_10 +   
## pool\_CloudResearch\_10 + pool\_Facebook\_10 + Trumppercent +   
## age, data = vi1\_m.data.cem)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.30770 -0.17352 -0.13557 -0.09352 0.93696   
##   
## Coefficients: (1 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 1.456e-01 1.633e-01 0.892 0.373  
## treated -4.134e-02 3.006e-02 -1.375 0.170  
## Female\_10 -4.209e-02 3.327e-02 -1.265 0.206  
## race\_white\_10 -5.724e-03 1.016e-01 -0.056 0.955  
## race\_black\_10 8.215e-02 1.102e-01 0.745 0.456  
## education\_low\_10 1.905e-02 4.907e-02 0.388 0.698  
## education\_medium\_10 -1.523e-02 4.906e-02 -0.310 0.756  
## pool\_CloudResearch\_10 2.474e-02 9.724e-02 0.254 0.799  
## pool\_Facebook\_10 NA NA NA NA  
## Trumppercent -1.033e-16 9.998e-17 -1.033 0.302  
## age 1.928e-03 1.409e-03 1.368 0.172  
##   
## Residual standard error: 0.3597 on 566 degrees of freedom  
## Multiple R-squared: 0.01795, Adjusted R-squared: 0.002331   
## F-statistic: 1.149 on 9 and 566 DF, p-value: 0.3257

### Genetic matching

set.seed(1)  
vi1\_m.out.gen <- matchit(treated ~ Female\_10 +  
 race\_white\_10 +  
 race\_black\_10 +  
 race\_other\_10 +  
 education\_low\_10 +  
 education\_medium\_10 +  
 education\_high\_10 +  
 pool\_CloudResearch\_10 +  
 pool\_Facebook\_10 +  
 pool\_Lucid\_10 +  
 Trumppercent +  
 age,   
 data = vi1\_data1,   
 replace = TRUE,   
 method = "genetic")  
vi1\_bal\_tab\_gen <- bal.tab(vi1\_m.out.gen, thresholds = c(m = 0.01), un = TRUE)  
vi1\_bal\_tab\_gen

## Balance Measures  
## Type Diff.Un Diff.Adj M.Threshold  
## distance Distance 0.2062 0.0064 Balanced, <0.01  
## Female\_10 Binary -0.0498 -0.0036 Balanced, <0.01  
## race\_white\_10 Binary -0.0211 -0.0018 Balanced, <0.01  
## race\_black\_10 Binary -0.0122 0.0018 Balanced, <0.01  
## race\_other\_10 Binary 0.0333 0.0000 Balanced, <0.01  
## education\_low\_10 Binary 0.0310 -0.0036 Balanced, <0.01  
## education\_medium\_10 Binary -0.0215 0.0000 Balanced, <0.01  
## education\_high\_10 Binary -0.0095 0.0036 Balanced, <0.01  
## pool\_CloudResearch\_10 Binary -0.0097 -0.0036 Balanced, <0.01  
## pool\_Facebook\_10 Binary 0.0120 0.0018 Balanced, <0.01  
## pool\_Lucid\_10 Binary -0.0023 0.0018 Balanced, <0.01  
## Trumppercent Contin. 0.0509 0.0054 Balanced, <0.01  
## age Contin. -0.0830 -0.0041 Balanced, <0.01  
##   
## Balance tally for mean differences  
## count  
## Balanced, <0.01 13  
## Not Balanced, >0.01 0  
##   
## Variable with the greatest mean difference  
## Variable Diff.Adj M.Threshold  
## Trumppercent 0.0054 Balanced, <0.01  
##   
## Sample sizes  
## Control Treated  
## All 513. 554  
## Matched (ESS) 205.71 554  
## Matched (Unweighted) 303. 554  
## Unmatched 210. 0

vi1\_lp.gen <- love.plot(  
 vi1\_m.out.gen,  
 threshold = .01,  
 binary = 'std',  
 treat = vi1\_data1$treated,  
 covs = vi1\_X\_1)  
  
vi1\_lp.gen



#### ATT by Gen

vi1\_m.data.gen <- match.data(vi1\_m.out.gen)  
head(vi1\_m.data.gen)

## # A tibble: 6 × 19  
## Treatment\_012 Female\_10 age\_young\_10 race\_white\_10 race\_black\_10 race\_other\_10  
## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 1 0 0 1 0 0  
## 2 1 0 1 1 0 0  
## 3 1 0 0 1 0 0  
## 4 1 1 1 0 1 0  
## 5 1 1 0 1 0 0  
## 6 0 0 1 0 1 0  
## # ℹ 13 more variables: education\_low\_10 <dbl>, education\_medium\_10 <dbl>,  
## # education\_high\_10 <dbl>, pool\_CloudResearch\_10 <dbl>,  
## # pool\_Facebook\_10 <dbl>, pool\_Lucid\_10 <dbl>, Trumppercent <dbl>,  
## # Trumphi <dbl>, age <dbl>, outcome <dbl>, treated <dbl>, distance <dbl>,  
## # weights <dbl>

#### Gen regression

vi1\_lm.gen1 <- lm(outcome ~ treated, data = vi1\_m.data.gen)  
summary(vi1\_lm.gen1)

##   
## Call:  
## lm(formula = outcome ~ treated, data = vi1\_m.data.gen)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.1782 -0.1782 -0.1354 -0.1354 0.8646   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.17822 0.02053 8.680 <2e-16 \*\*\*  
## treated -0.04284 0.02554 -1.677 0.0938 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3574 on 855 degrees of freedom  
## Multiple R-squared: 0.00328, Adjusted R-squared: 0.002114   
## F-statistic: 2.814 on 1 and 855 DF, p-value: 0.09382

vi1\_lm.gen2 <- lm(outcome ~ treated +   
 Female\_10 +  
 race\_white\_10 +  
 race\_black\_10 +  
 education\_low\_10 +  
 education\_medium\_10 +  
 pool\_CloudResearch\_10 +  
 pool\_Facebook\_10 +  
 Trumppercent +  
 age, data = vi1\_m.data.gen)  
summary(vi1\_lm.gen2)

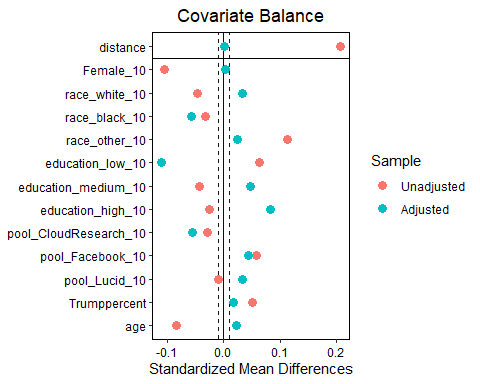
##   
## Call:  
## lm(formula = outcome ~ treated + Female\_10 + race\_white\_10 +   
## race\_black\_10 + education\_low\_10 + education\_medium\_10 +   
## pool\_CloudResearch\_10 + pool\_Facebook\_10 + Trumppercent +   
## age, data = vi1\_m.data.gen)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.53421 -0.16035 -0.12140 -0.08127 0.98515   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -5.520e-02 8.455e-02 -0.653 0.5140   
## treated -4.486e-02 2.503e-02 -1.792 0.0735 .   
## Female\_10 2.496e-03 2.558e-02 0.098 0.9223   
## race\_white\_10 7.208e-04 4.397e-02 0.016 0.9869   
## race\_black\_10 8.907e-02 5.025e-02 1.772 0.0767 .   
## education\_low\_10 2.936e-02 3.475e-02 0.845 0.3984   
## education\_medium\_10 1.509e-03 3.456e-02 0.044 0.9652   
## pool\_CloudResearch\_10 8.280e-02 4.474e-02 1.851 0.0645 .   
## pool\_Facebook\_10 4.136e-01 7.117e-02 5.812 8.75e-09 \*\*\*  
## Trumppercent 8.477e-17 7.526e-17 1.126 0.2603   
## age 1.961e-03 9.635e-04 2.036 0.0421 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3496 on 846 degrees of freedom  
## Multiple R-squared: 0.05654, Adjusted R-squared: 0.04539   
## F-statistic: 5.07 on 10 and 846 DF, p-value: 3.338e-07

### Nearest neighbour matching

vi1\_m.out.nnm <- matchit(treated ~ Female\_10 +  
 race\_white\_10 +  
 race\_black\_10 +  
 race\_other\_10 +  
 education\_low\_10 +  
 education\_medium\_10 +  
 education\_high\_10 +  
 pool\_CloudResearch\_10 +  
 pool\_Facebook\_10 +  
 pool\_Lucid\_10 +  
 Trumppercent +  
 age,   
 data = vi1\_data1,   
 replace = TRUE,   
 method = "nearest")  
vi1\_bal\_tab\_nnm <- bal.tab(vi1\_m.out.nnm, thresholds = c(m = 0.01), un = TRUE)  
vi1\_bal\_tab\_nnm

## Balance Measures  
## Type Diff.Un Diff.Adj M.Threshold  
## distance Distance 0.2062 0.0008 Balanced, <0.01  
## Female\_10 Binary -0.0498 0.0018 Balanced, <0.01  
## race\_white\_10 Binary -0.0211 0.0144 Not Balanced, >0.01  
## race\_black\_10 Binary -0.0122 -0.0217 Not Balanced, >0.01  
## race\_other\_10 Binary 0.0333 0.0072 Balanced, <0.01  
## education\_low\_10 Binary 0.0310 -0.0542 Not Balanced, >0.01  
## education\_medium\_10 Binary -0.0215 0.0235 Not Balanced, >0.01  
## education\_high\_10 Binary -0.0095 0.0307 Not Balanced, >0.01  
## pool\_CloudResearch\_10 Binary -0.0097 -0.0181 Not Balanced, >0.01  
## pool\_Facebook\_10 Binary 0.0120 0.0090 Balanced, <0.01  
## pool\_Lucid\_10 Binary -0.0023 0.0090 Balanced, <0.01  
## Trumppercent Contin. 0.0509 0.0169 Not Balanced, >0.01  
## age Contin. -0.0830 0.0219 Not Balanced, >0.01  
##   
## Balance tally for mean differences  
## count  
## Balanced, <0.01 5  
## Not Balanced, >0.01 8  
##   
## Variable with the greatest mean difference  
## Variable Diff.Adj M.Threshold  
## education\_low\_10 -0.0542 Not Balanced, >0.01  
##   
## Sample sizes  
## Control Treated  
## All 513. 554  
## Matched (ESS) 207.38 554  
## Matched (Unweighted) 289. 554  
## Unmatched 224. 0

vi1\_lp.nnm <- love.plot(  
 vi1\_m.out.nnm,  
 threshold = .01,  
 binary = 'std',  
 treat = vi1\_data1$treated,  
 covs = vi1\_X\_1)  
  
vi1\_lp.nnm



#### ATT by NNM

vi1\_m.data.nnm <- match.data(vi1\_m.out.nnm)  
head(vi1\_m.data.nnm)

## # A tibble: 6 × 19  
## Treatment\_012 Female\_10 age\_young\_10 race\_white\_10 race\_black\_10 race\_other\_10  
## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 1 0 0 1 0 0  
## 2 1 0 1 1 0 0  
## 3 1 0 0 1 0 0  
## 4 1 1 1 0 1 0  
## 5 1 1 0 1 0 0  
## 6 0 0 1 0 1 0  
## # ℹ 13 more variables: education\_low\_10 <dbl>, education\_medium\_10 <dbl>,  
## # education\_high\_10 <dbl>, pool\_CloudResearch\_10 <dbl>,  
## # pool\_Facebook\_10 <dbl>, pool\_Lucid\_10 <dbl>, Trumppercent <dbl>,  
## # Trumphi <dbl>, age <dbl>, outcome <dbl>, treated <dbl>, distance <dbl>,  
## # weights <dbl>

#### NNM regression

vi1\_lm.nnm1 <- lm(outcome ~ treated, data = vi1\_m.data.nnm)  
summary(vi1\_lm.nnm1)

##   
## Call:  
## lm(formula = outcome ~ treated, data = vi1\_m.data.nnm)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.1592 -0.1592 -0.1354 -0.1354 0.8646   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.15917 0.02064 7.712 3.49e-14 \*\*\*  
## treated -0.02379 0.02546 -0.934 0.35   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3509 on 841 degrees of freedom  
## Multiple R-squared: 0.001037, Adjusted R-squared: -0.0001506   
## F-statistic: 0.8732 on 1 and 841 DF, p-value: 0.3503

vi1\_lm.nnm2 <- lm(outcome ~ treated +   
 Female\_10 +  
 race\_white\_10 +  
 race\_black\_10 +  
 education\_low\_10 +  
 education\_medium\_10 +  
 pool\_CloudResearch\_10 +  
 pool\_Facebook\_10 +  
 Trumppercent +  
 age, data = vi1\_m.data.nnm)  
summary(vi1\_lm.nnm2)

##   
## Call:  
## lm(formula = outcome ~ treated + Female\_10 + race\_white\_10 +   
## race\_black\_10 + education\_low\_10 + education\_medium\_10 +   
## pool\_CloudResearch\_10 + pool\_Facebook\_10 + Trumppercent +   
## age, data = vi1\_m.data.nnm)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.50286 -0.15788 -0.11837 -0.06759 0.97892   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.038e-01 8.445e-02 -1.229 0.21939   
## treated -2.088e-02 2.496e-02 -0.836 0.40316   
## Female\_10 -2.969e-03 2.508e-02 -0.118 0.90582   
## race\_white\_10 8.506e-04 4.332e-02 0.020 0.98434   
## race\_black\_10 7.762e-02 4.920e-02 1.578 0.11498   
## education\_low\_10 7.338e-02 3.494e-02 2.100 0.03600 \*   
## education\_medium\_10 4.992e-02 3.478e-02 1.435 0.15157   
## pool\_CloudResearch\_10 7.199e-02 4.663e-02 1.544 0.12303   
## pool\_Facebook\_10 3.826e-01 7.276e-02 5.258 1.85e-07 \*\*\*  
## Trumppercent 1.590e-17 7.301e-17 0.218 0.82766   
## age 2.828e-03 9.215e-04 3.069 0.00222 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3431 on 832 degrees of freedom  
## Multiple R-squared: 0.0551, Adjusted R-squared: 0.04374   
## F-statistic: 4.852 on 10 and 832 DF, p-value: 8.052e-07

## Control vs Treatment 2 only (vi2)

### Experimental benchmark

Creating a basic and preliminary estimate of the treatment effect.

vi2\_cleaned <- vi\_cleaned %>%  
 filter(Treatment\_012 != 1)  
  
count(vi2\_cleaned, Treatment\_012)

## # A tibble: 2 × 2  
## Treatment\_012 n  
## <dbl> <int>  
## 1 0 516  
## 2 2 531

model\_vi2\_1 <- lm(outcome ~ treated, data = vi2\_cleaned)  
summary(model\_vi2\_1)

##   
## Call:  
## lm(formula = outcome ~ treated, data = vi2\_cleaned)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.2147 -0.2147 -0.1570 -0.1570 0.8430   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.15698 0.01711 9.176 <2e-16 \*\*\*  
## treated 0.05771 0.02402 2.402 0.0165 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3886 on 1045 degrees of freedom  
## Multiple R-squared: 0.005493, Adjusted R-squared: 0.004541   
## F-statistic: 5.772 on 1 and 1045 DF, p-value: 0.01646

Try adding other covariates into the regression and see

model\_vi2\_2 <- lm(outcome ~ treated +   
 Female\_10 +  
 race\_white\_10 +  
 race\_black\_10 +  
 education\_low\_10 +  
 education\_medium\_10 +  
 pool\_CloudResearch\_10 +  
 pool\_Facebook\_10 +  
 Trumppercent +  
 age, data = vi2\_cleaned)  
summary(model\_vi2\_2)

##   
## Call:  
## lm(formula = outcome ~ treated + Female\_10 + race\_white\_10 +   
## race\_black\_10 + education\_low\_10 + education\_medium\_10 +   
## pool\_CloudResearch\_10 + pool\_Facebook\_10 + Trumppercent +   
## age, data = vi2\_cleaned)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.53301 -0.20999 -0.16564 -0.06915 0.97689   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.091e-01 8.328e-02 -1.310 0.19047   
## treated 5.753e-02 2.382e-02 2.416 0.01588 \*   
## Female\_10 6.839e-03 2.570e-02 0.266 0.79023   
## race\_white\_10 -1.193e-02 4.838e-02 -0.246 0.80536   
## race\_black\_10 7.798e-03 5.420e-02 0.144 0.88563   
## education\_low\_10 2.436e-02 3.585e-02 0.679 0.49705   
## education\_medium\_10 -3.749e-02 3.500e-02 -1.071 0.28432   
## pool\_CloudResearch\_10 1.592e-01 4.437e-02 3.587 0.00035 \*\*\*  
## pool\_Facebook\_10 4.290e-01 7.926e-02 5.412 7.74e-08 \*\*\*  
## Trumppercent 7.456e-17 7.205e-17 1.035 0.30097   
## age 2.238e-03 8.897e-04 2.515 0.01206 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3828 on 1031 degrees of freedom  
## (5 observations deleted due to missingness)  
## Multiple R-squared: 0.04329, Adjusted R-squared: 0.03401   
## F-statistic: 4.665 on 10 and 1031 DF, p-value: 1.555e-06

### Matching preparation

vi2\_tr <- which(vi2\_cleaned$Treatment\_012 == 2)  
vi2\_ct <- which(vi2\_cleaned$Treatment\_012 == 0)  
vi2\_ntr <- length(vi2\_tr)  
vi2\_ntr

## [1] 531

vi2\_nct <- length(vi2\_ct)  
vi2\_nct

## [1] 516

### Basic estimate

mean(na.omit(vi2\_cleaned$outcome[vi2\_tr])) - mean(na.omit(vi2\_cleaned$outcome[vi2\_ct]))

## [1] 0.05771252

### Covariate balance

vi2\_covs1 <- vi2\_cleaned %>%  
 dplyr::select(Female\_10,  
 race\_white\_10,  
 race\_black\_10,  
 race\_other\_10,  
 education\_low\_10,  
 education\_medium\_10,  
 education\_high\_10,  
 pool\_CloudResearch\_10,  
 pool\_Facebook\_10,  
 pool\_Lucid\_10,  
 Trumppercent,  
 age,  
 treated) %>%  
 na.omit()  
  
bal.tab(vi2\_covs1, treat = vi2\_covs1$treated)

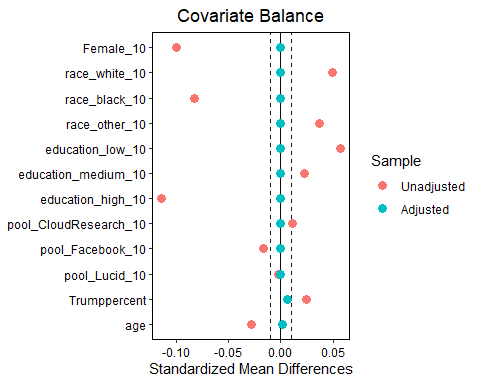
## Balance Measures  
## Type Diff.Un  
## Female\_10 Binary -0.0476  
## race\_white\_10 Binary 0.0208  
## race\_black\_10 Binary -0.0303  
## race\_other\_10 Binary 0.0095  
## education\_low\_10 Binary 0.0280  
## education\_medium\_10 Binary 0.0114  
## education\_high\_10 Binary -0.0394  
## pool\_CloudResearch\_10 Binary 0.0034  
## pool\_Facebook\_10 Binary -0.0029  
## pool\_Lucid\_10 Binary -0.0005  
## Trumppercent Contin. 0.0235  
## age Contin. -0.0278  
## treated Binary 1.0000  
##   
## Sample sizes  
## Control Treated  
## All 513 529

### CEM

vi2\_data1 <- na.omit(vi2\_cleaned)  
  
vi2\_X\_1 <- vi2\_data1 %>%  
 dplyr::select(Female\_10,  
 race\_white\_10,  
 race\_black\_10,  
 race\_other\_10,  
 education\_low\_10,  
 education\_medium\_10,  
 education\_high\_10,  
 pool\_CloudResearch\_10,  
 pool\_Facebook\_10,  
 pool\_Lucid\_10,  
 Trumppercent,  
 age)  
  
vi2\_m.out.cem <- matchit(treated ~ Female\_10 +  
 race\_white\_10 +  
 race\_black\_10 +  
 race\_other\_10 +  
 education\_low\_10 +  
 education\_medium\_10 +  
 education\_high\_10 +  
 pool\_CloudResearch\_10 +  
 pool\_Facebook\_10 +  
 pool\_Lucid\_10 +  
 Trumppercent +  
 age,  
 data = vi2\_data1,  
 method = "cem")  
  
vi2\_bal\_tab\_cem <- bal.tab(vi2\_m.out.cem,  
 thresholds = c(m = 0.01),  
 un = TRUE)  
vi2\_bal\_tab\_cem

## Balance Measures  
## Type Diff.Un Diff.Adj M.Threshold  
## Female\_10 Binary -0.0476 0.0000 Balanced, <0.01  
## race\_white\_10 Binary 0.0208 0.0000 Balanced, <0.01  
## race\_black\_10 Binary -0.0303 0.0000 Balanced, <0.01  
## race\_other\_10 Binary 0.0095 0.0000 Balanced, <0.01  
## education\_low\_10 Binary 0.0280 0.0000 Balanced, <0.01  
## education\_medium\_10 Binary 0.0114 0.0000 Balanced, <0.01  
## education\_high\_10 Binary -0.0394 0.0000 Balanced, <0.01  
## pool\_CloudResearch\_10 Binary 0.0034 0.0000 Balanced, <0.01  
## pool\_Facebook\_10 Binary -0.0029 0.0000 Balanced, <0.01  
## pool\_Lucid\_10 Binary -0.0005 0.0000 Balanced, <0.01  
## Trumppercent Contin. 0.0243 0.0063 Balanced, <0.01  
## age Contin. -0.0285 0.0017 Balanced, <0.01  
##   
## Balance tally for mean differences  
## count  
## Balanced, <0.01 12  
## Not Balanced, >0.01 0  
##   
## Variable with the greatest mean difference  
## Variable Diff.Adj M.Threshold  
## Trumppercent 0.0063 Balanced, <0.01  
##   
## Sample sizes  
## Control Treated  
## All 513. 529  
## Matched (ESS) 183.64 307  
## Matched (Unweighted) 284. 307  
## Unmatched 229. 222

vi2\_lp.cem <- love.plot(  
 vi2\_m.out.cem,  
 threshold = .01,  
 binary = 'std',  
 treat = vi2\_data1$treated,  
 covs = vi2\_X\_1)  
  
vi2\_lp.cem



#### ATT by CEM

vi2\_m.data.cem <- match.data(vi2\_m.out.cem)  
head(vi2\_m.data.cem)

## # A tibble: 6 × 19  
## Treatment\_012 Female\_10 age\_young\_10 race\_white\_10 race\_black\_10 race\_other\_10  
## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 2 1 0 1 0 0  
## 2 0 0 1 0 1 0  
## 3 2 0 1 1 0 0  
## 4 0 1 1 1 0 0  
## 5 2 0 1 1 0 0  
## 6 0 1 1 1 0 0  
## # ℹ 13 more variables: education\_low\_10 <dbl>, education\_medium\_10 <dbl>,  
## # education\_high\_10 <dbl>, pool\_CloudResearch\_10 <dbl>,  
## # pool\_Facebook\_10 <dbl>, pool\_Lucid\_10 <dbl>, Trumppercent <dbl>,  
## # Trumphi <dbl>, age <dbl>, outcome <dbl>, treated <dbl>, weights <dbl>,  
## # subclass <fct>

#### CEM regression

vi2\_lm.cem1 <- lm(outcome ~ treated, data = vi2\_m.data.cem)  
summary(vi2\_lm.cem1)

##   
## Call:  
## lm(formula = outcome ~ treated, data = vi2\_m.data.cem)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.2117 -0.2117 -0.1373 -0.1373 0.8627   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.13732 0.02253 6.096 1.97e-09 \*\*\*  
## treated 0.07440 0.03125 2.381 0.0176 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3796 on 589 degrees of freedom  
## Multiple R-squared: 0.009529, Adjusted R-squared: 0.007848   
## F-statistic: 5.667 on 1 and 589 DF, p-value: 0.01761

vi2\_lm.cem2 <- lm(outcome ~ treated +   
 Female\_10 +  
 race\_white\_10 +  
 race\_black\_10 +  
 education\_low\_10 +  
 education\_medium\_10 +  
 pool\_CloudResearch\_10 +  
 pool\_Facebook\_10 +  
 Trumppercent +  
 age, data = vi2\_m.data.cem)  
summary(vi2\_lm.cem2)

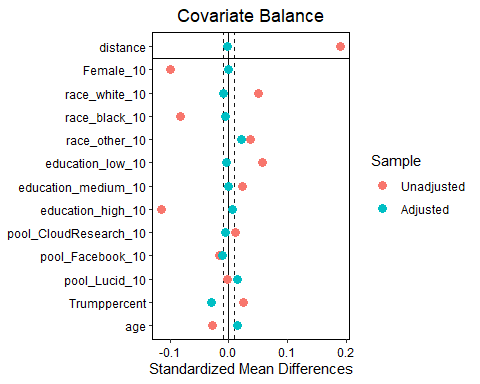
##   
## Call:  
## lm(formula = outcome ~ treated + Female\_10 + race\_white\_10 +   
## race\_black\_10 + education\_low\_10 + education\_medium\_10 +   
## pool\_CloudResearch\_10 + pool\_Facebook\_10 + Trumppercent +   
## age, data = vi2\_m.data.cem)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.48913 -0.20683 -0.15520 -0.07314 0.95473   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 8.491e-02 2.236e-01 0.380 0.7043   
## treated 7.215e-02 3.106e-02 2.323 0.0205 \*  
## Female\_10 7.969e-03 3.570e-02 0.223 0.8234   
## race\_white\_10 -3.548e-01 1.912e-01 -1.856 0.0640 .  
## race\_black\_10 -3.348e-01 1.961e-01 -1.707 0.0883 .  
## education\_low\_10 9.758e-02 5.814e-02 1.679 0.0938 .  
## education\_medium\_10 3.936e-02 5.760e-02 0.683 0.4946   
## pool\_CloudResearch\_10 1.494e-01 9.342e-02 1.599 0.1103   
## pool\_Facebook\_10 -6.366e-02 2.813e-01 -0.226 0.8210   
## Trumppercent 1.431e-16 1.047e-16 1.367 0.1722   
## age 3.328e-03 1.335e-03 2.494 0.0129 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3763 on 580 degrees of freedom  
## Multiple R-squared: 0.04171, Adjusted R-squared: 0.02519   
## F-statistic: 2.525 on 10 and 580 DF, p-value: 0.005574

### Genetic matching

set.seed(1)  
vi2\_m.out.gen <- matchit(treated ~ Female\_10 +  
 race\_white\_10 +  
 race\_black\_10 +  
 race\_other\_10 +  
 education\_low\_10 +  
 education\_medium\_10 +  
 education\_high\_10 +  
 pool\_CloudResearch\_10 +  
 pool\_Facebook\_10 +  
 pool\_Lucid\_10 +  
 Trumppercent +  
 age,   
 data = vi2\_data1,   
 replace = TRUE,   
 method = "genetic")  
vi2\_bal\_tab\_gen <- bal.tab(vi2\_m.out.gen, thresholds = c(m = 0.01), un = TRUE)  
vi2\_bal\_tab\_gen

## Balance Measures  
## Type Diff.Un Diff.Adj M.Threshold  
## distance Distance 0.1892 -0.0019 Balanced, <0.01  
## Female\_10 Binary -0.0476 0.0000 Balanced, <0.01  
## race\_white\_10 Binary 0.0208 -0.0038 Balanced, <0.01  
## race\_black\_10 Binary -0.0303 -0.0019 Balanced, <0.01  
## race\_other\_10 Binary 0.0095 0.0057 Balanced, <0.01  
## education\_low\_10 Binary 0.0280 -0.0019 Balanced, <0.01  
## education\_medium\_10 Binary 0.0114 0.0000 Balanced, <0.01  
## education\_high\_10 Binary -0.0394 0.0019 Balanced, <0.01  
## pool\_CloudResearch\_10 Binary 0.0034 -0.0019 Balanced, <0.01  
## pool\_Facebook\_10 Binary -0.0029 -0.0019 Balanced, <0.01  
## pool\_Lucid\_10 Binary -0.0005 0.0038 Balanced, <0.01  
## Trumppercent Contin. 0.0243 -0.0302 Not Balanced, >0.01  
## age Contin. -0.0285 0.0139 Not Balanced, >0.01  
##   
## Balance tally for mean differences  
## count  
## Balanced, <0.01 11  
## Not Balanced, >0.01 2  
##   
## Variable with the greatest mean difference  
## Variable Diff.Adj M.Threshold  
## Trumppercent -0.0302 Not Balanced, >0.01  
##   
## Sample sizes  
## Control Treated  
## All 513. 529  
## Matched (ESS) 201.18 529  
## Matched (Unweighted) 287. 529  
## Unmatched 226. 0

vi2\_lp.gen <- love.plot(  
 vi2\_m.out.gen,  
 threshold = .01,  
 binary = 'std',  
 treat = vi2\_data1$treated,  
 covs = vi2\_X\_1)  
  
vi2\_lp.gen



#### ATT by Gen

vi2\_m.data.gen <- match.data(vi2\_m.out.gen)  
head(vi2\_m.data.gen)

## # A tibble: 6 × 19  
## Treatment\_012 Female\_10 age\_young\_10 race\_white\_10 race\_black\_10 race\_other\_10  
## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 2 0 1 1 0 0  
## 2 2 1 0 1 0 0  
## 3 2 0 0 1 0 0  
## 4 0 0 1 0 1 0  
## 5 2 1 1 1 0 0  
## 6 2 0 1 1 0 0  
## # ℹ 13 more variables: education\_low\_10 <dbl>, education\_medium\_10 <dbl>,  
## # education\_high\_10 <dbl>, pool\_CloudResearch\_10 <dbl>,  
## # pool\_Facebook\_10 <dbl>, pool\_Lucid\_10 <dbl>, Trumppercent <dbl>,  
## # Trumphi <dbl>, age <dbl>, outcome <dbl>, treated <dbl>, distance <dbl>,  
## # weights <dbl>

#### Gen regression

vi2\_lm.gen1 <- lm(outcome ~ treated, data = vi2\_m.data.gen)  
summary(vi2\_lm.gen1)

##   
## Call:  
## lm(formula = outcome ~ treated, data = vi2\_m.data.gen)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.2136 -0.2136 -0.2136 -0.1429 0.8571   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.14286 0.02304 6.201 8.94e-10 \*\*\*  
## treated 0.07075 0.02861 2.473 0.0136 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3903 on 814 degrees of freedom  
## Multiple R-squared: 0.007455, Adjusted R-squared: 0.006236   
## F-statistic: 6.114 on 1 and 814 DF, p-value: 0.01361

vi2\_lm.gen2 <- lm(outcome ~ treated +   
 Female\_10 +  
 race\_white\_10 +  
 race\_black\_10 +  
 education\_low\_10 +  
 education\_medium\_10 +  
 pool\_CloudResearch\_10 +  
 pool\_Facebook\_10 +  
 Trumppercent +  
 age, data = vi2\_m.data.gen)  
summary(vi2\_lm.gen2)

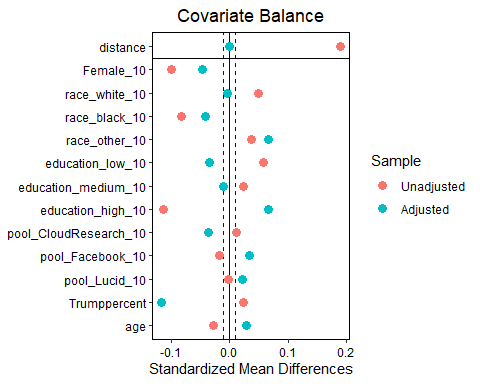
##   
## Call:  
## lm(formula = outcome ~ treated + Female\_10 + race\_white\_10 +   
## race\_black\_10 + education\_low\_10 + education\_medium\_10 +   
## pool\_CloudResearch\_10 + pool\_Facebook\_10 + Trumppercent +   
## age, data = vi2\_m.data.gen)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.47045 -0.21898 -0.16793 -0.05906 0.92566   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.698e-01 9.558e-02 -1.776 0.076063 .   
## treated 6.820e-02 2.831e-02 2.409 0.016212 \*   
## Female\_10 1.305e-02 2.880e-02 0.453 0.650589   
## race\_white\_10 4.646e-03 5.403e-02 0.086 0.931497   
## race\_black\_10 1.645e-02 6.144e-02 0.268 0.788970   
## education\_low\_10 1.903e-02 4.218e-02 0.451 0.651987   
## education\_medium\_10 -4.396e-02 4.114e-02 -1.069 0.285603   
## pool\_CloudResearch\_10 1.745e-01 5.079e-02 3.436 0.000621 \*\*\*  
## pool\_Facebook\_10 3.737e-01 9.443e-02 3.958 8.24e-05 \*\*\*  
## Trumppercent 1.054e-16 8.421e-17 1.252 0.211018   
## age 2.526e-03 1.032e-03 2.448 0.014570 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3856 on 805 degrees of freedom  
## Multiple R-squared: 0.04207, Adjusted R-squared: 0.03017   
## F-statistic: 3.536 on 10 and 805 DF, p-value: 0.0001365

### Nearest neighbour matching

vi2\_m.out.nnm <- matchit(treated ~ Female\_10 +  
 race\_white\_10 +  
 race\_black\_10 +  
 race\_other\_10 +  
 education\_low\_10 +  
 education\_medium\_10 +  
 education\_high\_10 +  
 pool\_CloudResearch\_10 +  
 pool\_Facebook\_10 +  
 pool\_Lucid\_10 +  
 Trumppercent +  
 age,   
 data = vi2\_data1,   
 replace = TRUE,   
 method = "nearest")  
vi2\_bal\_tab\_nnm <- bal.tab(vi2\_m.out.nnm, thresholds = c(m = 0.01), un = TRUE)  
vi2\_bal\_tab\_nnm

## Balance Measures  
## Type Diff.Un Diff.Adj M.Threshold  
## distance Distance 0.1892 -0.0001 Balanced, <0.01  
## Female\_10 Binary -0.0476 -0.0227 Not Balanced, >0.01  
## race\_white\_10 Binary 0.0208 -0.0019 Balanced, <0.01  
## race\_black\_10 Binary -0.0303 -0.0151 Not Balanced, >0.01  
## race\_other\_10 Binary 0.0095 0.0170 Not Balanced, >0.01  
## education\_low\_10 Binary 0.0280 -0.0170 Not Balanced, >0.01  
## education\_medium\_10 Binary 0.0114 -0.0057 Balanced, <0.01  
## education\_high\_10 Binary -0.0394 0.0227 Not Balanced, >0.01  
## pool\_CloudResearch\_10 Binary 0.0034 -0.0113 Not Balanced, >0.01  
## pool\_Facebook\_10 Binary -0.0029 0.0057 Balanced, <0.01  
## pool\_Lucid\_10 Binary -0.0005 0.0057 Balanced, <0.01  
## Trumppercent Contin. 0.0243 -0.1166 Not Balanced, >0.01  
## age Contin. -0.0285 0.0285 Not Balanced, >0.01  
##   
## Balance tally for mean differences  
## count  
## Balanced, <0.01 5  
## Not Balanced, >0.01 8  
##   
## Variable with the greatest mean difference  
## Variable Diff.Adj M.Threshold  
## Trumppercent -0.1166 Not Balanced, >0.01  
##   
## Sample sizes  
## Control Treated  
## All 513. 529  
## Matched (ESS) 199.74 529  
## Matched (Unweighted) 286. 529  
## Unmatched 227. 0

vi2\_lp.nnm <- love.plot(  
 vi2\_m.out.nnm,  
 threshold = .01,  
 binary = 'std',  
 treat = vi2\_data1$treated,  
 covs = vi2\_X\_1)  
  
vi2\_lp.nnm



#### ATT by NNM

vi2\_m.data.nnm <- match.data(vi2\_m.out.nnm)  
head(vi2\_m.data.nnm)

## # A tibble: 6 × 19  
## Treatment\_012 Female\_10 age\_young\_10 race\_white\_10 race\_black\_10 race\_other\_10  
## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 2 0 1 1 0 0  
## 2 2 1 0 1 0 0  
## 3 2 0 0 1 0 0  
## 4 0 0 1 0 1 0  
## 5 2 1 1 1 0 0  
## 6 0 0 0 1 0 0  
## # ℹ 13 more variables: education\_low\_10 <dbl>, education\_medium\_10 <dbl>,  
## # education\_high\_10 <dbl>, pool\_CloudResearch\_10 <dbl>,  
## # pool\_Facebook\_10 <dbl>, pool\_Lucid\_10 <dbl>, Trumppercent <dbl>,  
## # Trumphi <dbl>, age <dbl>, outcome <dbl>, treated <dbl>, distance <dbl>,  
## # weights <dbl>

#### NNM regression

vi2\_lm.nnm1 <- lm(outcome ~ treated, data = vi2\_m.data.nnm)  
summary(vi2\_lm.nnm1)

##   
## Call:  
## lm(formula = outcome ~ treated, data = vi2\_m.data.nnm)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.2136 -0.2136 -0.2136 -0.1434 0.8566   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.14336 0.02309 6.208 8.54e-10 \*\*\*  
## treated 0.07025 0.02866 2.451 0.0145 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3905 on 813 degrees of freedom  
## Multiple R-squared: 0.007336, Adjusted R-squared: 0.006115   
## F-statistic: 6.008 on 1 and 813 DF, p-value: 0.01445

vi2\_lm.nnm2 <- lm(outcome ~ treated +   
 Female\_10 +  
 race\_white\_10 +  
 race\_black\_10 +  
 education\_low\_10 +  
 education\_medium\_10 +  
 pool\_CloudResearch\_10 +  
 pool\_Facebook\_10 +  
 Trumppercent +  
 age, data = vi2\_m.data.nnm)  
summary(vi2\_lm.nnm2)

##   
## Call:  
## lm(formula = outcome ~ treated + Female\_10 + race\_white\_10 +   
## race\_black\_10 + education\_low\_10 + education\_medium\_10 +   
## pool\_CloudResearch\_10 + pool\_Facebook\_10 + Trumppercent +   
## age, data = vi2\_m.data.nnm)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.47810 -0.21597 -0.16957 -0.06068 0.94161   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.743e-01 9.689e-02 -1.799 0.072417 .   
## treated 7.352e-02 2.836e-02 2.592 0.009714 \*\*   
## Female\_10 1.676e-02 2.918e-02 0.574 0.565984   
## race\_white\_10 -3.700e-03 5.354e-02 -0.069 0.944921   
## race\_black\_10 4.692e-03 6.088e-02 0.077 0.938596   
## education\_low\_10 5.910e-02 4.227e-02 1.398 0.162501   
## education\_medium\_10 -1.821e-02 4.145e-02 -0.439 0.660619   
## pool\_CloudResearch\_10 1.767e-01 5.195e-02 3.402 0.000701 \*\*\*  
## pool\_Facebook\_10 3.862e-01 9.193e-02 4.201 2.96e-05 \*\*\*  
## Trumppercent 9.099e-17 8.290e-17 1.098 0.272738   
## age 2.040e-03 1.039e-03 1.962 0.050066 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3856 on 804 degrees of freedom  
## Multiple R-squared: 0.043, Adjusted R-squared: 0.0311   
## F-statistic: 3.613 on 10 and 804 DF, p-value: 0.0001019

# To-do:

matching

covariate balance

summary statistics

comparison between the datasets